

# COUPLINGS, DISTANCES AND CONTRACTIVITY FOR DIFFUSION PROCESSES REVISITED

ANDREAS EBERLE

**ABSTRACT.** We consider contractivity for diffusion semigroups w.r.t. Kantorovich ( $L^1$  Wasserstein) distances based on appropriately chosen concave functions. These distances are inbetween total variation and usual Wasserstein distances. It is shown that by appropriate explicit choices of the underlying distance, contractivity with rates of close to optimal order can be obtained in several fundamental classes of examples where contractivity w.r.t. standard Wasserstein distances fails. Applications include overdamped Langevin diffusions with locally non-convex potentials, products of these processes, and systems of weakly interacting diffusions.

## 1. INTRODUCTION

Consider a diffusion process  $(X_t)_{t \geq 0}$  in  $\mathbb{R}^d$  defined by a stochastic differential equation

$$(1.1) \quad dX_t = b(X_t) dt + \sigma dB_t.$$

Here  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion,  $\sigma \in \mathbb{R}^{d \times d}$  is a constant  $d \times d$  matrix with  $\det \sigma > 0$ , and  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a locally Lipschitz continuous function. We assume that the unique strong solution of (1.1) is non-explosive, which is essentially a consequence of the assumptions imposed further below. The transition kernels of the diffusion process on  $\mathbb{R}^d$  defined by (1.1) will be denoted by  $p_t(x, dy)$ .

Contraction properties of the transition semigroup  $(p_t)_{t \geq 0}$  have been studied by various approaches. In particular,  $L^2$  and entropy methods yield bounds that both are relatively stable under perturbations and applicable in high dimensions, cf. e.g. [2, 3, 1, 32, 36, 4]. On the other hand, coupling methods provide a more intuitive probabilistic understanding of convergence to equilibrium [30, 29, 11, 10, 13, 36, 34]. In contrast to  $L^2$  and entropy methods, bounds resulting from coupling methods typically hold for arbitrary initial values  $x_0 \in \mathbb{R}^d$ . In many applications, couplings are used to bound the total variation distances  $d_{TV}(\mu p_t, \nu p_t)$  between the laws  $\mu p_t$  and  $\nu p_t$  of  $X_t$  w.r.t. two different initial distributions  $\mu$  and  $\nu$  at a given time  $t \geq 0$ , cf. [30, 29]. Typically, however, the total variation distance is decaying substantially only after a certain amount of time, and the map  $\mu \mapsto \mu p_t$  is not a contraction for small times  $t$ . This is also manifested in cut-off phenomena [15, 28, 16, 9].

Alternatively, it is well-known that basic couplings (i.e., couplings given by the flow of the s.d.e. (1.1)) can be used to show that the map  $\mu \mapsto \mu p_t$  is exponentially

contractive w.r.t.  $L^p$  Wasserstein distances  $W^p$  for any  $p \in [1, \infty)$  if, for example,  $(X_t)$  is an overdamped Langevin diffusion with a strictly convex potential  $U \in C^2(\mathbb{R}^d)$ , i.e.,  $\sigma = I_d$  and  $b = -\nabla U/2$ . This leads to an elegant and powerful approach to convergence to equilibrium and to many related results if applicable. However, it has been pointed out in [35] that strict convexity of  $U$  is also a necessary condition for exponential contractivity w.r.t.  $W^p$ . This seems to limit the applicability substantially.

Here, we are instead considering exponential contractivity w.r.t. Kantorovich ( $L^1$  Wasserstein) distances  $W_f$  based on underlying distance functions of the form

$$d_f(x, y) = f(\|x - y\|) \quad \text{on } \mathbb{R}^d,$$

and, more generally,

$$d_f(x, y) = \sum_{i=1}^n f_i(\|x^i - y^i\|) \quad \text{on } \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n},$$

where  $f, f_i : [0, \infty) \rightarrow [0, \infty)$  are strictly increasing *concave* functions, cf. Sections 2.1 and 3.1 below for details. For proving exponential contractivity, we will apply a reflection coupling on  $\mathbb{R}^d$  and an (approximate) componentwise reflection coupling on products of Euclidean spaces. It will become clear by the proofs below, that for distances based on concave functions  $f, f_i$ , these couplings are superior to basic couplings, whereas the basic couplings are superior w.r.t. the Wasserstein distances  $W^p$  for  $p > 1$ , cf. e.g. Lemma 6.2.

The idea to study contraction properties w.r.t. Kantorovich distances based on concave distance functions goes back to Chen and Wang [12, 13], where it is implicitly contained in the proofs. Indeed, in [13] and [36], Chen and Wang apply very similar methods to estimate spectral gaps of diffusion generators on  $\mathbb{R}^d$  and on manifolds. Related arguments have also been applied by Hairer and Mattingly in [20] to quantify exponential ergodicity in infinite dimensional situations. The key idea in the results below is to “almost” optimize the choice of the functions  $f$  and  $f_i$  to obtain a large exponential decay rate. In the case  $n = 1$ , this idea has been exploited intensively by Chen and Wang in [13] to derive lower bounds for spectral gaps. The novelty here is that we suggest a simple and very *explicit* choice for  $f$  that leads to close to optimal results in several examples. The extension to the product case based on an approximate componentwise reflection coupling then enables us to obtain dimension free contraction results in product models and perturbations thereof without relying on convexity.

**Example 1.1 (Overdamped Langevin dynamics with locally non-convex potential).** Suppose that  $\sigma = I_d$  and  $b(x) = -\frac{1}{2}\nabla U(x)$  for a function  $U \in C^2(\mathbb{R}^d)$  that is *strictly convex outside a given ball*  $B \subset \mathbb{R}^d$ . Then  $Z := \int \exp(-U(x))dx$  is finite, and the probability measure

$$d\mu = Z^{-1} \exp(-U) dx$$

is a stationary distribution for the diffusion process  $(X_t)$ . Corollary 2.3 below yields exponential contractivity for the transition semigroup  $(p_t)$  with an explicit

rate w.r.t. an appropriate Kantorovich distance  $W_f$ . As a consequence, we obtain dimension-independent upper bounds for the standard  $L^1$  Wasserstein distances between the laws  $\nu p_t$  of  $X_t$  and  $\mu$  for arbitrary initial distributions  $\nu$  and  $t \geq 0$ . These bounds are of optimal order in  $R, L \in [0, \infty)$  and  $K \in (0, \infty)$  if  $(x - y) \cdot (\nabla U(x) - \nabla U(y))$  is bounded from below by  $-L|x - y|^2$  for  $|x - y| < R$  and by  $K|x - y|^2$  for  $|x - y| \geq R$ .

**Example 1.2 (Product models).** For a diffusion process  $X_t = (X_t^1, \dots, X_t^n)$  in  $\mathbb{R}^{n \cdot d}$  with independent Langevin diffusions  $X^1, \dots, X^n$  as in Example 1.1, Theorem 3.1 below yields exponential contractivity in an appropriate Kantorovich distance with rate  $c = \min(c_1, \dots, c_n)$  where  $c_1, \dots, c_n$  are the lower bounds obtained for the contraction rates of the components.

**Example 1.3 (Systems of interacting diffusions).** More generally, consider a system

$$dX_t^i = -\frac{1}{2}\nabla U(X_t^i) dt - \frac{\alpha}{n} \sum_{j=1}^n \nabla V(X_t^i - X_t^j) dt + dB_t^i, \quad i = 1, \dots, n,$$

of  $n$  interacting diffusion processes in  $\mathbb{R}^d$  where  $U \in C^2(\mathbb{R}^d)$  is strictly convex outside a ball,  $V \in C^2(\mathbb{R}^d)$  has bounded second derivatives, and  $B^1, \dots, B^n$  are independent Brownian motions in  $\mathbb{R}^d$ . Then Corollary 3.4 below shows that for  $\alpha$  sufficiently small, exponential contractivity holds in an appropriate Kantorovich distance with a rate that does not depend on  $n$ .

Before stating the results in detail, we briefly introduce the couplings to be considered in the proofs below:

A *coupling by reflection* of two solutions of (1.1) with initial distributions  $\mu$  and  $\nu$  is a diffusion process  $(X_t, Y_t)$  with values in  $\mathbb{R}^{2d}$  defined by  $(X_0, Y_0) \sim \eta$  where  $\eta$  is a coupling of  $\mu$  and  $\nu$ ,

$$\begin{aligned} dX_t &= b(X_t) dt + \sigma dB_t && \text{for } t \geq 0, \\ (1.2) \quad dY_t &= b(Y_t) dt + \sigma(I - 2e_t e_t^\top) dB_t && \text{for } t < T, \quad Y_t = X_t \text{ for } t \geq T. \end{aligned}$$

Here  $e_t e_t^\top$  is the orthogonal projection onto the unit vector

$$e_t := \sigma^{-1}(X_t - Y_t) / |\sigma^{-1}(X_t - Y_t)|,$$

and  $T = \inf\{t \geq 0 : X_t = Y_t\}$  is the coupling time, i.e., the first hitting time of the diagonal  $\Delta = \{(x, y) \in \mathbb{R}^{2d} : x = y\}$ , cf. [30, 11]. The reflection coupling can be realized as a diffusion process in  $\mathbb{R}^{2d}$ , and the marginal processes  $(X_t)_{t \geq 0}$  and  $(Y_t)_{t \geq 0}$  are solutions of (1.1) w.r.t. the Brownian motions  $B_t$  and

$$\check{B}_t = \int_0^t (I_d - 2I_{\{s < T\}} e_s e_s^\top) dB_s.$$

Notice that by Lévy's characterization,  $\check{B}$  is indeed a Brownian motion since the process  $I_d - 2I_{\{s < T\}} e_s e_s^\top$  takes values in the orthogonal matrices. The difference vector

$$Z_t := X_t - Y_t$$

solves the s.d.e.

$$(1.3) \quad \begin{aligned} dZ_t &= (b(X_t) - b(Y_t)) dt + 2|\sigma^{-1}Z_t|^{-1}Z_t dW_t \quad \text{for } t < T, \\ Z_t &= 0 \quad \text{for } t \geq T, \end{aligned}$$

w.r.t. the *one-dimensional* Brownian motion

$$W_t = \int_0^t e_s^\top dB_s.$$

A *basic coupling* of two solutions of (1.1) is defined correspondingly with  $e_t \equiv 0$ , i.e., the same noise is applied both to  $X_t$  and  $Y_t$ . Below we will also consider *mixed couplings* that are reflection couplings for certain values of  $Z_t$ , basic couplings for other values of  $Z_t$ , and mixtures of both types of couplings for  $Z_t$  in an intermediate region. Notice that the standard reflection coupling introduced above is a basic coupling for  $t \geq T$ , i.e., if  $Z_t = 0$  !

More generally, we will consider couplings for diffusion processes on product spaces (such as in Examples 1.2 and 1.3) that are approximately *componentwise reflection couplings*, i.e., the  $i$ -th component  $(X_t^i, Y_t^i)$  of the coupling  $(X_t, Y_t)$  is defined similarly to (1.2) provided  $|X_t^i - Y_t^i| \geq \delta$  for a given constant  $\delta > 0$ , cf. Section 6 below.

For diffusion processes with non-constant diffusion matrix  $\sigma(x)$ , the reflection coupling should be replaced by the *Kendall-Cranston coupling* w.r.t. the intrinsic Riemannian metric  $G(x) = (\sigma(x)\sigma(x)^\top)^{-1}$  induced by the diffusion coefficients, cf. [26, 14, 23, 36]. Here, we restrict ourselves to the case of constant diffusion matrices where the Kendall-Cranston coupling coincides with the standard coupling by reflection.

In the next two sections we state the main results of this paper. A part of the results in Section 2 have been announced in the Comptes Rendus Note [17].

## 2. MAIN RESULTS FOR REFLECTION COUPLING

**2.1. Reflection couplings and contractivity on  $\mathbb{R}^d$ .** Lindvall and Rogers [30] introduced coupling by reflection in order to derive upper bounds for the total variation distance of the distributions of  $X_t$  and  $Y_t$  at a given time  $t \geq 0$ . Here we are instead considering the Kantorovich-Rubinstein ( $L^1$ -Wasserstein) distances

$$(2.1) \quad W_f(\mu, \nu) = \inf_{\eta} \int d_f(x, y) \eta(dx dy), \quad d_f(x, y) = f(\|x - y\|) \quad (x, y \in \mathbb{R}^d),$$

of probability measures  $\mu, \nu$  on  $\mathbb{R}^d$ , where the infimum is over all couplings  $\eta$  of  $\mu$  and  $\nu$ ,  $f : [0, \infty) \rightarrow [0, \infty)$  is an appropriately chosen concave increasing function with  $f(0) = 0$ , and  $\|z\| = \sqrt{z \cdot Gz}$  with  $G \in \mathbb{R}^{d \times d}$  symmetric and strictly positive definite. Typical choices for the norm are the Euclidean norm  $\|z\| = |z|$  and the intrinsic metric  $\|z\| = |\sigma^{-1}z|$  corresponding to  $G = I_d$  and  $G = (\sigma\sigma^\top)^{-1}$  respectively.

**Remark 2.1 (Interpolating between total variation and Wasserstein distances).** For the choice of the function  $f$  there are two extreme cases with minimal and maximal concavity:

- (1) Choosing  $f(x) = x$  yields the *standard Kantorovich ( $L^1$  Wasserstein) distance*  $W_f = W^1$ . In this case it is well known that if, for example,  $G = \sigma = I_d$  and  $b(x) = -\nabla U(x)/2$ , then the transition kernels  $p_t(x, dy)$  of the diffusion process  $(X_t)$  satisfy

$$W_f(\mu p_t, \nu p_t) \leq e^{-Kt/2} W_f(\mu, \nu) \quad \text{for any } \mu, \nu \text{ and } t \geq 0,$$

provided  $\nabla^2 U \geq K \cdot I_d$  holds globally. This condition is also sharp in the sense that if  $U$  is not globally strictly convex, then contractivity of  $p_t$  w.r.t.  $W_f$  does not hold, cf. Sturm and von Renesse [35].

- (2) On the other hand, choosing  $f(x) = I_{(0,\infty)}(x)$  yields the *total variation distance*  $W_f = d_{TV}$ . In this case,

$$W_f(\mu p_t, \nu p_t) \leq \mathbb{P}[T > t] \quad \text{for any } \mu, \nu \text{ and } t \geq 0,$$

but there is no contractivity of  $p_t$  w.r.t.  $d_{TV}$  in general. Indeed, in many applications  $d_{TV}(\mu p_t, \nu p_t)$  only decreases substantially after a certain amount of time (“cut-off phenomenon”).

Following Chen and Wang [13], we will argue that by choosing for  $f$  an appropriate concave function, exponential contractivity w.r.t.  $W_f$  holds even without global convexity. We now describe how the function  $f$  can be chosen in a very explicit way such that the obtained exponential decay rate w.r.t. the Kantorovich distance  $W_f$  differs from the maximal decay rate that we can achieve by our approach based on reflection coupling only by a constant factor.

At first, similarly to Lindvall and Rogers [30], let us define for  $r \in (0, \infty)$ :

$$\kappa(r) = \inf \left\{ -2 \frac{|\sigma^{-1}(x-y)|^2}{\|x-y\|^2} \frac{(x-y) \cdot G(b(x) - b(y))}{\|x-y\|^2} : x, y \in \mathbb{R}^d \text{ s.t. } \|x-y\| = r \right\},$$

i.e.,  $\kappa(r)$  is the largest constant such that

$$(2.2) \quad (x-y) \cdot G(b(x) - b(y)) \leq -\frac{1}{2} \kappa(r) \|x-y\|^4 / |\sigma^{-1}(x-y)|^2$$

holds for any  $x, y \in \mathbb{R}^d$  with  $\|x-y\| = r$ . Notice that if  $\|\cdot\|$  is the intrinsic metric then the factor  $|\sigma^{-1}(x-y)|^2 / \|x-y\|^2$  equals 1. In Example 1.1 with  $G = I_d$ , we have

$$\kappa(r) = \inf \left\{ \int_0^1 \partial_{(x-y)/|x-y|}^2 U((1-t)x + ty) dt : x, y \in \mathbb{R}^d \text{ s.t. } |x-y| = r \right\}.$$

We assume from now on that  $\kappa(r)$  is a continuous function on  $(0, \infty)$  satisfying

$$(2.3) \quad \liminf_{r \rightarrow \infty} \kappa(r) > 0 \quad \text{and} \quad \int_0^1 r \kappa(r)^- dr < \infty.$$

In Example 1.1 with  $G = I_d$ , this assumption is satisfied if  $U$  is strictly convex outside a ball.

Next, we define constants  $R_0, R_1 \in [0, \infty)$  with  $R_0 \leq R_1$  by

$$(2.4) \quad R_0 = \inf\{R \geq 0 : \kappa(r) \geq 0 \ \forall r \geq R\},$$

$$(2.5) \quad R_1 = \inf\{R \geq R_0 : \kappa(r)R(R - R_0) \geq 8 \ \forall r \geq R\},$$

Notice that by (2.3), both constants are finite. We now consider the particular distance function  $d_f(x, y) = f(\|x - y\|)$  given by

$$(2.6) \quad f(r) = \int_0^r \varphi(s)g(s) \, ds, \quad \text{where}$$

$$\varphi(r) = \exp\left(-\frac{1}{4} \int_0^r s\kappa(s)^- \, ds\right), \quad \Phi(r) = \int_0^r \varphi(s) \, ds,$$

$$g(r) = 1 - \frac{1}{2} \frac{\int_0^{r \wedge R_1} \frac{\Phi(s)}{\varphi(s)} \, ds}{\int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} \, ds}.$$

Let us summarize some basic properties of the functions  $\varphi, g$  and  $f$ :

- $\varphi$  is decreasing,  $\varphi(0) = 1$ , and  $\varphi(r) = \varphi(R_0)$  for any  $r \geq R_0$ ,
- $g$  is decreasing,  $g(0) = 1$ , and  $g(r) = \frac{1}{2}$  for any  $r \geq R_1$ ,
- $f$  is concave,  $f(0) = 0$ ,  $f'(0) = 1$ , and

$$(2.7) \quad \Phi(r)/2 \leq f(r) \leq \Phi(r) \quad \text{for any } r \geq 0.$$

The last statement shows that  $d_f$  and  $d_\Phi$  as well as  $W_f$  and  $W_\Phi$  differ at most by a factor 2.

We will explain in Section 4 below how the choice of  $f$  is obtained by trying to maximize the exponential decay rate. Let us now state our first main result which will also be proven in Section 4.

**Theorem 2.2 (Exponential contractivity of reflection coupling).** *Let  $\alpha := \sup\{|\sigma^{-1}z|^2 : z \in \mathbb{R}^d \text{ with } \|z\| = 1\}$ , and define  $c \in (0, \infty)$  by*

$$(2.8) \quad \frac{1}{c} = \alpha \int_0^{R_1} \Phi(s)\varphi(s)^{-1} \, ds = \alpha \int_0^{R_1} \int_0^s \exp\left(\frac{1}{4} \int_t^s u\kappa(u)^- \, du\right) \, dt \, ds.$$

*Then for the distance  $d_f$  given by (2.1) and (2.6), the function  $t \mapsto e^{ct}\mathbb{E}[d_f(X_t, Y_t)]$  is decreasing on  $[0, \infty)$ .*

The theorem yields exponential contractivity at rate  $c > 0$  for the transition kernels  $p_t$  of (1.1) w.r.t. the Kantorovich distance  $W_f$ . Moreover, it implies upper bounds for the standard Kantorovich ( $L^1$  Wasserstein) distance  $W^1 = W_{\text{id}}$  w.r.t. the distance function  $d(x, y) = \|x - y\|$ :

**Corollary 2.3.** *For any  $t \geq 0$  and any probability measures  $\mu, \nu$  on  $\mathbb{R}^d$ ,*

$$(2.9) \quad W_f(\mu p_t, \nu p_t) \leq \exp(-ct) W_f(\mu, \nu), \quad \text{and}$$

$$(2.10) \quad W^1(\mu p_t, \nu p_t) \leq 2\varphi(R_0)^{-1} \exp(-ct) W^1(\mu, \nu).$$

Note that the second estimate follows from the first, since by the properties of  $\varphi$  and  $g$  stated above,  $\varphi(R_0)/2 \leq f' \leq 1$ , and hence

$$(2.11) \quad \varphi(R_0)\|x - y\|/2 \leq d_f(x, y) \leq \|x - y\| \quad \text{for any } x, y \in \mathbb{R}^d.$$

The corollary yields an upper bound for mixing times w.r.t. the Kantorovich distance  $W^1$ . For  $\varepsilon > 0$  let

$$\tau_{W^1}(\varepsilon) := \inf\{t \geq 0 : W^1(\mu p_t, \nu p_t) \leq \varepsilon W^1(\mu, \nu) \forall \mu, \nu \in \mathcal{M}_1(\mathbb{R}^d)\}.$$

Then by Corollary 2.3,

$$\tau_{W^1}(\varepsilon) \leq c^{-1} \log(2/(\varepsilon \varphi(R_0))) \quad \text{for any } \varepsilon > 0.$$

The proofs of Theorem 2.2 and Corollary 2.3 are given in Section 4 below.

**Remark 2.4 (Non-constant diffusion coefficients).** The methods and results presented above have natural extensions to diffusion processes with non-constant diffusion matrices. In that case, one possibility is to use an ad hoc coupling as in [30], but this leads to restrictive assumptions and bounds that are from optimal. A better approach is to switch to a Riemannian setup where the metric is the intrinsic metric  $G(x) = (\sigma(x)\sigma(x)^T)^{-1}$  given by the diffusion coefficients. Then by replacing the Reflection Coupling by the corresponding Kendall-Cranston coupling, one should expect similar results as above.

**2.2. Consequences.** We summarize some important consequences of exponential contractivity w.r.t. Kantorovich distances as stated in Corollary 2.3. These consequences are essentially well-known, cf. e.g. Joulin and Ollivier [25], Joulin [24] and Komorowski and Walczuk [27] for related results in discrete and continuous time respectively. For the reader's convenience, the proofs are nevertheless included in Section 4 below.

We assume from now on that  $\|z\| = |\sigma^{-1}z|$  is the intrinsic metric,  $b$  is in  $C^1(\mathbb{R}^d, \mathbb{R}^d)$ , and

$$(2.12) \quad \int |z| p_t(x_0, dz) < \infty$$

holds for some  $x_0 \in \mathbb{R}^d$  and any  $t \geq 0$ . Then, equivalently to (2.9), Theorem 2.2 implies Lipschitz contractivity for the transition semigroup

$$(p_t g)(x) = \int g(z) p_t(x, dz)$$

w.r.t. the metric  $d_f$ , i.e.,

$$(2.13) \quad \|p_t g\|_{\text{Lip}(f)} \leq \exp(-ct) \|g\|_{\text{Lip}(f)}$$

holds for any  $t \geq 0$  and any Lipschitz continuous function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ , where

$$\|g\|_{\text{Lip}(f)} = \sup \left\{ \frac{|g(x) - g(y)|}{d_f(x, y)} : x, y \in \mathbb{R}^d \text{ s.t. } x \neq y \right\}$$

denotes the Lipschitz semi-norm w.r.t.  $d_f$ . An immediate consequence is the existence of a unique stationary distribution  $\mu$  with finite second moments:



**Corollary 2.5 (Convergence to equilibrium).** *There exists a unique stationary distribution  $\mu$  of  $(p_t)_{t \geq 0}$  satisfying  $\int |y| \mu(dy) < \infty$ , and*

$$(2.14) \quad \text{Var}_\mu(g) \leq (2c)^{-1} \|g\|_{\text{Lip}(f)}^2 \quad \text{for any Lipschitz continuous } g : \mathbb{R}^d \rightarrow \mathbb{R}.$$

Moreover, for any probability measure  $\nu$  on  $\mathbb{R}^d$ ,

$$(2.15) \quad W_f(\mu, \nu p_t) \leq \exp(-ct) W_f(\mu, \nu) \quad \text{for any } t \geq 0.$$

We refer to [5, 8] for other recent results on convergence to equilibrium of diffusion processes in Wasserstein distances.

Further important consequences of (2.13) are quantitative non-asymptotic bounds for the decay of correlations and the bias and variance of ergodic averages. Let  $x_0 \in \mathbb{R}^d$  and suppose that  $(X, \mathbb{P})$  is a solution of (1.1) with initial condition  $X_0 = x_0$  a.s.

**Corollary 2.6 (Decay of correlations).** *For any Lipschitz continuous functions  $g, h : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $s, t \geq 0$ ,*

$$\text{Cov}(g(X_t), h(X_{t+s})) \leq \frac{1 - e^{-2ct}}{2c} e^{-cs} \|g\|_{\text{Lip}(f)} \|h\|_{\text{Lip}(f)}.$$

**Corollary 2.7 (Bias and variance of ergodic averages).** *For any Lipschitz continuous function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $t \in (0, \infty)$ ,*

$$\begin{aligned} \left| \mathbb{E} \left[ \frac{1}{t} \int_0^t g(X_s) ds - \int g d\mu \right] \right| &\leq \frac{1 - e^{-ct}}{ct} \|g\|_{\text{Lip}(f)} \int d_f(x_0, y) \mu(dy), \quad \text{and} \\ \text{Var} \left( \frac{1}{t} \int_0^t g(X_s) ds \right) &\leq \frac{1}{c^2 t} \|g\|_{\text{Lip}(f)}^2. \end{aligned}$$

In the variance estimate in Corollary 2.7, one of the factors  $1/c$  is due to the variance bound (2.14) w.r.t. the stationary distribution whereas the second factor  $1/c$  bounds the decay rate for the correlations. Short proofs of Corollaries 2.5, 2.6, and 2.7 are included in Section 4.

**Remark 2.8 (Central Limit Theorem, Gaussian deviation inequality).** The contractivity w.r.t.  $W_f$  can also be used to prove a central limit theorem for the ergodic averages [27] and a Gaussian deviation inequality strengthening Corollary 2.7, cf. Remark 2.10 in [24].

**2.3. First examples.** In order to illustrate the bounds given in Theorem 2.2 and in Corollary 2.3, we estimate the constant  $c$  defined by (2.8) in different scenarios, and we study the behaviour of  $c$  under perturbations of the drift  $b$ .

We first consider the situation where  $\kappa$  is bounded from below by a negative constant for any  $r$ , and by a positive constant for large  $r$ :



**Lemma 2.9 (Contractivity under lower bounds on  $\kappa$ ).** *Suppose that*

$$(2.16) \quad \kappa(r) \geq -L \text{ for } r \leq R, \text{ and } \kappa(r) \geq K \text{ for } r > R$$

*hold with constants  $R, L \in [0, \infty)$  and  $K \in (0, \infty)$ . If  $LR_0^2 \leq 8$  then*

$$(2.17) \quad \alpha^{-1}c^{-1} \leq \frac{e-1}{2}R^2 + e\sqrt{8K^{-1}}R + 4K^{-1} \leq \frac{3e}{2} \max(R^2, 8K^{-1}),$$

*and if  $LR_0^2 \geq 8$  then*

$$(2.18) \quad \alpha^{-1}c^{-1} \leq 8\sqrt{2\pi}R^{-1}L^{-1/2}(L^{-1} + K^{-1}) \exp\left(\frac{LR^2}{8}\right) + 32R^{-2}K^{-2}.$$

**Remark 2.10 (Convex case).** If  $L = 0$  then the bound in (2.17) improves to

$$(2.19) \quad \alpha^{-1}c^{-1} \leq 2 \max(R^2, 2K^{-1}).$$

The proofs of Lemma 2.9 and Remark 2.10 are given in Section 5 below.

In the first case considered in the lemma, the constant  $c$  is at least of order  $\min(R^{-2}, K)$ . Even if  $L = 0$  (convex case), this order can not be improved as one-dimensional Langevin diffusions with potential  $U(x) = Kx^2/2$ , or, respectively, with vanishing drift on  $(-R/2, R/2)$  demonstrate. In particular, for  $U(x) = Kx^2/2$  with  $K > 0$ , the distance  $W_f$  is equivalent to  $W^1$ , and the exact decay rate is  $K/2$ . This differs from the bounds in (2.19) and (2.17) only by a factor 2, 6e respectively. Thus if  $LR_0^2$  is not too large, the contractivity properties are not affected substantially by non-convexity.

In the second case ( $LR_0^2 \geq 8$ ), if  $K \geq \text{const.} \cdot L$  then the upper bound for  $c^{-1}$  is of order  $L^{-3/2}R^{-1} \exp(LR^2/8)$ . This order in  $R$  and  $L$  is again optimal:

**Example 2.11 (Double-well potential with  $U''(x) = -L$  for  $|x| \leq R/2$ ).** Consider a Langevin diffusion in  $\mathbb{R}^1$  with a symmetric potential  $U \in C^2(\mathbb{R})$  satisfying  $U(x) = -Lx^2/2$  for  $x \in [-R/2, R/2]$ ,  $U'' \geq -L$ , and  $\liminf_{|x| \rightarrow \infty} U''(x) > 0$ . If  $\|\cdot\|$  is the Euclidean norm then  $\kappa(r) = -L$  for  $r \in (0, R]$ . On the other hand, let  $\tau_0 = \inf\{t \geq 0 : X_t = 0\}$  denote the first hitting time of 0. Then for any initial condition  $x_0 > 0$ ,

$$(2.20) \quad \lim_{t \rightarrow \infty} t^{-1} \log P_{x_0}[\tau_0 > t] = -\lambda_1(0, \infty)$$

where  $-\lambda_1(0, \infty)$  is the first Dirichlet eigenvalue of the generator  $\mathcal{L}v = (v'' - U'v)/2$  on  $(0, \infty)$ , cf. [19] or see Section 5 below for a short proof of the corresponding lower bound that is relevant here. If  $LR^2 \geq 4$  then by inserting the function  $g(x) = \min(\sqrt{L}x, 1)$  into the variational characterization of the Dirichlet eigenvalue, we obtain the upper bound

$$(2.21) \quad \lambda_1(0, \infty) \leq \frac{3}{4}e^{1/2}L^{3/2}R \exp(-LR^2/8),$$

cf. Section 5 below. The estimates (2.20) and (2.21) seem to indicate that for  $x_0 > 0$ , the Kantorovich distance  $W^1(\delta_{-x_0}p_t, \delta_{x_0}p_t)$  decays at most with a rate

of order  $L^{3/2}R \exp(-LR^2/8)$ . Indeed, under appropriate growth assumptions on  $U(x)$  for  $|x| \geq R$ , one can prove that

$$\mathbb{P}_R[\tau_0 > t] \geq 3/4 \quad \text{for any } t \leq \lambda_1(0, \infty)^{-1}/4,$$

cf. Section 5. Hence for  $t \leq 3^{-1}e^{-1/2}L^{-3/2}R^{-1} \exp(LR^2/8)$ , the Kantorovich distance  $W^1(\delta_{Rp_t}, \mu)$  between  $\delta_{Rp_t}$  and the stationary distribution  $\mu$  is bounded from below by a strictly positive constant that does not depend on  $L$  and  $R$  if  $LR^2 \geq 4$ .

For analysing the behaviour of  $c$  under perturbations of the drift, we assume that  $\|z\| = |\sigma^{-1}z|$  is the intrinsic metric corresponding to the diffusion matrix, i.e.,  $G = (\sigma\sigma^T)^{-1}$ . Suppose that

$$(2.22) \quad b(x) = b_0(x) + \gamma(x) \quad \text{for any } x \in \mathbb{R}$$

with locally Lipschitz continuous functions  $b_0, \gamma : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . For  $r > 0$  let

$$(2.23) \quad \kappa_0(r) = \inf \left\{ -2 \frac{(x-y) \cdot G(b_0(x) - b_0(y))}{\|x-y\|^2} : x, y \in \mathbb{R}^d \text{ s.t. } \|x-y\| = r \right\}$$

be defined analogously to  $\kappa(r)$  with  $b$  replaced by  $b_0$ . We assume that  $\kappa_0$  satisfies the assumptions (2.3) imposed on  $\kappa$  above, and we define  $R_0$  and  $R_1$  similarly to (2.4) and (2.5) but with  $\kappa$  replaced by  $\kappa_0$ . Now suppose that there exists a constant  $R \leq R_0$  such that

$$(2.24) \quad (x-y) \cdot (\gamma(x) - \gamma(y)) \leq 0 \quad \text{for any } x, y \in \mathbb{R}^d \text{ s.t. } \|x-y\| \geq R.$$

Then  $\kappa(r) \geq \kappa_0(r)$  for  $r \geq R$ , and hence the constants  $R_0$  and  $R_1$  defined w.r.t.  $b$  are smaller than the corresponding constants defined w.r.t.  $b_0$ . In this situation, we can compare the lower bounds  $c$  and  $c_0$  for the contraction rates w.r.t.  $b$  and  $b_0$  given by (2.8):

**Lemma 2.12 (Bounded and Lipschitz perturbations).** *Suppose that the drift  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is given by (2.22) with  $b_0$  and  $\gamma$  satisfying the assumptions stated above, and let  $c$  and  $c_0$  denote the lower bounds for the contraction rates w.r.t.  $b$  and  $b_0$  given by (2.8).*

(1) *If  $\gamma$  is bounded and (2.24) holds for a constant  $R \in [0, R_0]$  then*

$$(2.25) \quad c \geq c_0 \exp(-R \sup \|\gamma\|).$$

(2) *If  $\gamma$  satisfies the one-sided Lipschitz condition*

$$(2.26) \quad (x-y) \cdot G(\gamma(x) - \gamma(y)) \leq L \cdot \|x-y\|^2 \quad \forall x, y \in \mathbb{R}^d$$

*with a finite constant  $L \in [0, \infty)$  and (2.24) holds for a constant  $R \in [0, R_0]$  then*

$$(2.27) \quad c \geq c_0 \exp(-LR^2/4).$$

**Remark 2.13.** The condition  $R \leq R_0$  is required in Lemma 2.12. If (2.24) does not hold for  $x, y \in \mathbb{R}^d$  with  $\|x-y\| \geq R_0$  then the constants  $R_0(b)$  and  $R_1(b)$  defined w.r.t.  $b$  are in general greater than the corresponding constants defined w.r.t.  $b_0$ , i.e., the region of non-convexity increases by adding the drift  $\gamma$ . This will also affect the bound in (2.8) significantly.

The proof of Lemma 2.12 is given in Section 5.

**2.4. Local contractivity and a high-dimensional example.** Consider again the setup in Section 2.1. In some applications, the condition  $\liminf_{r \rightarrow \infty} \kappa(r) > 0$  imposed above is not satisfied, but the diffusion process will stay inside a ball  $B \subset \mathbb{R}^d$  for a long time with high probability. In this case, one can still prove exponential contractivity up to an error term that is determined by the exit probabilities from the ball. Corresponding estimates are useful to prove non-asymptotic error bounds, i.e., for fixed  $t \in (0, \infty)$ , cf. e.g. [7, 6, 18].

Fix  $R \in (0, \infty)$  and let  $W_{f_R}$  denote the Kantorovich distance based on the distance function  $d_{f_R}(x, y) = f_R(\|x - y\|)$  given by

$$(2.28) \quad f_R(r) = \int_0^r \varphi(s) g_R(s) ds \quad \text{for } r \geq 0,$$

where  $\varphi$  and  $\Phi$  are defined by (2.6), and

$$(2.29) \quad g_R(r) = 1 - \int_0^{r \wedge R} \frac{\Phi(s)}{\varphi(s)} ds \Big/ \int_0^R \frac{\Phi(s)}{\varphi(s)} ds.$$

Notice that

$$g_R(r) = 0 \quad \text{and} \quad f_R(r) = f_R(R) \quad \text{for any } r \geq R,$$

i.e., we have cut the distance at  $f_R(R)$ .

**Theorem 2.14 (Local exponential contractivity).** *Suppose that the assumptions from Section 2.1 are satisfied except for the condition  $\liminf_{r \rightarrow \infty} \kappa(r) > 0$ . Then for any  $t, R \geq 0$  and any probability measures  $\mu, \nu$  on  $\mathbb{R}^d$ ,*

$$(2.30) \quad \begin{aligned} W_{f_R}(\mu p_t, \nu p_t) &\leq \exp(-c_R t) W_{f_R}(\mu, \nu) \\ &\quad + R \cdot (\mathbb{P}_\mu[\tau_{R/2} \leq t] + \mathbb{P}_\nu[\tau_{R/2} \leq t]), \end{aligned}$$

where  $(X_t, \mathbb{P}_\mu)$  is a diffusion process satisfying (1.1) with initial distribution  $\mu$ ,  $\tau_{R/2} = \inf\{t \geq 0 : \|X_t\| > R/2\}$  denotes the first exit time from the ball of radius  $R/2$  around 0, and

$$(2.31) \quad \frac{1}{c_R} = \alpha \int_0^R \Phi(s) \varphi(s)^{-1} ds = \alpha \int_0^R \int_0^s \exp\left(\frac{1}{4} \int_t^s u \kappa(u)^- du\right) dt ds.$$

The proof of the theorem is given in Section 5. In applications, the exit probabilities are typically estimated by using appropriate Lyapunov functions.

**Example 2.15 (Stochastic heat equation).** We consider the diffusion in  $\mathbb{R}^{d-1}$  given by  $X_t^0 \equiv X_t^d \equiv 0$  and

$$(2.32) \quad dX_t^i = [d^2(X_t^{i+1} - 2X_t^i + X_t^{i-1}) + V'(X_t^i)] dt + \sqrt{d} dB_t^i,$$

$i = 1, \dots, d-1$ , where  $V : \mathbb{R} \rightarrow \mathbb{R}$  is a  $C^2$  function such that  $V'' \geq -L$  for a finite constant  $L \in \mathbb{R}$ . The equation (2.32) is a spatial discretization at the grid points

$i/d$  ( $i = 0, 1, \dots, d$ ) of the stochastic heat equation with space-time white noise and Dirichlet boundary conditions on the interval  $[0, 1]$  given by

$$(2.33) \quad du = \Delta_{\text{Dir}} u + V'(u) dt + dW$$

with the Dirichlet Laplacian  $\Delta_{\text{Dir}}$  on the interval  $[0, 1]$  and a cylindrical Wiener process  $(W_t)_{t \geq 0}$  over the Hilbert space  $L^2(0, 1)$ . We observe that (2.32) is of the form (1.1) with  $\sigma = \sqrt{d}I_{d-1}$  and  $b = -d\nabla U$  where

$$U(x) = \frac{d}{2} \sum_{i=1}^d |x^i - x^{i-1}|^2 + \frac{1}{d} \sum_{i=0}^d V(x^i)$$

for  $x = (x^1, \dots, x^{d-1}) \in \mathbb{R}^{d-1}$  and  $x^0 = x^d = 0$ . By the discrete Poincaré inequality,

$$\sum_{i=1}^d |x^i - x^{i-1}|^2 \geq 2(1 - \cos(\pi/d)) \sum_{i=1}^{d-1} |x^i|^2.$$

Hence for any  $x, \xi \in \mathbb{R}^{d-1}$  and  $x^0 = x^d = \xi^0 = \xi^d = 0$ , the lower bound

$$\partial_{\xi\xi}^2 U(x) = d \sum_{i=1}^d |\xi^i - \xi^{i-1}|^2 + \frac{1}{d} \sum_{i=1}^{d-1} V''(x^i) |\xi^i|^2 \geq \frac{1}{d} K_d \sum_{i=1}^{d-1} |\xi^i|^2$$

holds with  $K_d = 2d^2(1 - \cos(\pi/d)) - L$ , and thus

$$(x - y) \cdot (b(x) - b(y)) = -d(x - y) \cdot (\nabla U(x) - \nabla U(y)) \leq -K_d |x - y|^2$$

for any  $x, y \in \mathbb{R}^{d-1}$  where  $|\cdot|$  denotes the Euclidean norm. Choosing for  $\|\cdot\|$  the intrinsic metric  $\|x\| = d^{-1/2}|x|$ , we obtain

$$\kappa(r) \geq 2K_d \quad \text{for any } r > 0.$$

In particular, the function  $\kappa$  is bounded from below uniformly by a real constant that does not depend on the dimension  $d$  since

$$(2.34) \quad \lim_{d \rightarrow \infty} K_d = \pi^2 - L > -\infty.$$

Theorem 2.14 now shows that for any  $R > 0$ , local exponential contractivity in the sense of (2.30) holds on the ball

$$B_{R/2} = \{x \in \mathbb{R}^{d-1} : \|x\| \leq R/2\} = \{x \in \mathbb{R}^{d-1} : |x| \leq d^{1/2}R/2\}$$

with rate  $c_R$  satisfying

$$\begin{aligned} \frac{1}{c_R} &\leq 4\sqrt{\pi}R^{-1}|K_d|^{-3/2} \exp(-K_d R^2/4) && \text{for } K_d R^2 \leq -4, \\ \frac{1}{c_R} &\leq (e-1)R^2/2 && \text{for } -4 \leq K_d R^2 < 0, \\ \frac{1}{c_R} &\leq R^2/2 && \text{for } K_d = 0 \text{ respectively.} \end{aligned}$$

Here the explicit upper bounds are obtained analogously as in the proof of Lemma 2.9. For  $K_d > 0$ , strict convexity holds, and we obtain global exponential contractivity with a dimension-independent rate. We remark that because of (2.34),

the bounds also carry over to the limiting SPDE (2.33) for which they imply local exponential contractivity on balls w.r.t. the  $L^2$  norm.

### 3. MAIN RESULTS ON PRODUCT SPACES

**3.1. Componentwise reflection coupling and contractivity on product spaces.** We now consider a system

$$(3.1) \quad dX_t^i = b^i(X_t) dt + dB_t^i, \quad i = 1, \dots, n,$$

of  $n$  interacting diffusion processes taking values in  $\mathbb{R}^{d_i}$ ,  $d_i \in \mathbb{N}$ . Here  $B^i$ ,  $i = 1, \dots, n$ , are independent Brownian motions in  $\mathbb{R}^{d_i}$ ,  $X = (X^1, \dots, X^n)$  is a diffusion process taking values in  $\mathbb{R}^d$  where  $d = \sum_{i=1}^n d_i$ , and  $b^i : \mathbb{R}^d \rightarrow \mathbb{R}^{d_i}$  are locally Lipschitz continuous functions. We will assume that

$$(3.2) \quad b^i(x) = b_0^i(x^i) + \gamma^i(x), \quad i = 1, \dots, n,$$

where the functions  $b_0^i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d_i}$  are locally Lipschitz continuous, and  $\gamma^i : \mathbb{R}^d \rightarrow \mathbb{R}^{d_i}$  are “sufficiently small” perturbations, cf. Theorem 3.1 below. In particular, for  $\gamma^i \equiv 0$  the components  $X^1, \dots, X^n$  are independent.

To analyse contraction properties of the process  $X$ , one could use a reflection coupling on  $\mathbb{R}^d$  and apply the results above based on a distance function of the form  $d_f(x, y) = f(|x - y|)$ . In some applications, this approach does indeed provide dimension-free bounds, cf. Example 2.15 above. However, in the product case  $\gamma^i \equiv 0$  it leads in general to lower bounds for contraction rates that degenerate rapidly as  $n \rightarrow \infty$ , even though one would expect exponential contractivity with the minimum of the contraction rates for the components. The reason is that the approach requires convexity outside a Euclidean ball in  $\mathbb{R}^d$  whereas in corresponding product models, in general convexity only holds if all components are outside given balls in  $\mathbb{R}^{d_i}$ .

Instead, we now consider contractivity w.r.t. Kantorovich distances  $W_{f,w}$  based on distance functions on  $\mathbb{R}^d = \mathbb{R}^{d_1+\dots+d_n}$  of the form

$$(3.3) \quad d_{f,w}(x, y) = \sum_{i=1}^n f_i(|x^i - y^i|) w_i.$$

Here  $f_i : [0, \infty) \rightarrow [0, \infty)$ ,  $1 \leq i \leq n$ , are strictly increasing concave  $C^1$  functions with  $f_i(0) = 0$  and  $f_i'(0) = 1$  that are obtained from  $b_0^i$  in the same way as  $f$  has been obtained from  $b$  above, and  $w_i \in (0, 1]$  are positive weights. In many applications, one can choose  $w_i = 1$  for any  $i$ . The corresponding distance will then be denoted by  $d_{1,f}$ . Notice that  $d_{1,f}$  is bounded from above by the  $\ell^1$  distance

$$d_{\ell^1}(x, y) = \sum_{i=1}^n |x^i - y^i|.$$

Hence  $W_{1,f}$  is bounded from above by the Kantorovich distance  $W_{\ell^1}$  based on  $d_{\ell^1}$ .

For  $r \in (0, \infty)$  let

$$(3.4) \quad \kappa_i(r) = r^{-2} \inf \{ -2(x - y) \cdot (b_0^i(x) - b_0^i(y)) : x, y \in \mathbb{R}^{d_i} \text{ s.t. } |x - y| = r \}.$$

Similarly as above, we assume that for  $1 \leq i \leq n$ ,

$$(3.5) \quad \kappa_i : (0, \infty) \rightarrow \mathbb{R} \text{ is continuous with } \liminf_{r \rightarrow \infty} \kappa_i(r) > 0.$$

Moreover, we assume

$$(3.6) \quad \lim_{r \rightarrow 0} r \kappa_i(r) = 0.$$

Let  $R_0^i$ ,  $R_1^i$ ,  $g_i(r)$ ,  $\varphi_i(r)$ ,  $f_i(r)$  and  $\Phi_i(r) = \int_0^r \varphi_i(s) ds$  be defined analogously to (2.4), (2.5) and (2.6) with  $\kappa$  replaced by  $\kappa_i$ . Moreover, we define  $c_i \in (0, \infty)$  by

$$(3.7) \quad \frac{1}{c_i} = \int_0^{R_1^i} \Phi_i(s) \varphi_i(s)^{-1} ds = \int_0^{R_1^i} \int_0^s \exp\left(\frac{1}{4} \int_t^s u \kappa_i(u)^{-} du\right) dt ds.$$

Recall that by Theorem 2.2 and Corollary 2.3,  $c_i$  is a lower bound for the contraction rate of the diffusion process  $\tilde{X}^i$  on  $\mathbb{R}^{d_i}$  satisfying the s.d.e.  $d\tilde{X}_t^i = b_0^i(\tilde{X}_t^i) + dB_t^i$ .

Let  $p_t(x, dy)$  denote the transition kernels of the diffusion process  $X_t = (X_t^1, \dots, X_t^d)$  on  $\mathbb{R}^d$  satisfying (3.1). We now state our second main result:

**Theorem 3.1 (Exponential contractivity on product spaces).** *Suppose that (3.5) and (3.6) hold, and suppose that there exist constants  $\varepsilon_i \in [0, c_i)$ ,  $1 \leq i \leq n$ , such that for any  $x, y \in \mathbb{R}^d$ ,*

$$(3.8) \quad \sum_{i=1}^n |\gamma^i(x) - \gamma^i(y)| w_i \leq \sum_{i=1}^n \varepsilon_i f_i(|x^i - y^i|) w_i.$$

Then for any  $t \geq 0$  and any probability measures  $\mu, \nu$  on  $\mathbb{R}^d$ ,

$$(3.9) \quad W_{f,w}(\mu p_t, \nu p_t) \leq \exp(-ct) W_{f,w}(\mu, \nu), \quad \text{and}$$

$$(3.10) \quad W_{\ell^1}(\mu p_t, \nu p_t) \leq A \exp(-ct) W_{\ell^1}(\mu, \nu),$$

where  $c = \min_{i=1, \dots, n} (c_i - \varepsilon_i)$  and  $A = 2 \bigg/ \min_{i=1, \dots, n} (\varphi_i(R_0^i) w_i)$ .

**Example 3.2 (Product model).** In the product case,  $\gamma^i \equiv 0$  for any  $i$ . Hence Condition (3.8) is satisfied with  $\varepsilon_i = 0$ , and, therefore,

$$W_{f,w}(\mu p_t, \nu p_t) \leq \exp(-ct) W_{f,w}(\mu, \nu)$$

holds with  $c = \min c_i$  for any choice of the weights  $w_1, \dots, w_n$ .

More generally than in the example, suppose now that  $\gamma = (\gamma^1, \dots, \gamma^n)$  satisfies an  $\ell^1$ -Lipschitz condition

$$(3.11) \quad \sum_{i=1}^n |\gamma^i(x) - \gamma^i(y)| \leq \lambda \sum_{i=1}^n |x^i - y^i| \quad \forall x, y \in \mathbb{R}^d.$$

Then exponential contractivity holds for the perturbed product model provided  $\lambda < c_i \varphi(R_0^i)/2$  for any  $i$ :

**Corollary 3.3 (Perturbations of product models).** *Suppose that (3.2), (3.5), (3.6) and (3.11) hold with  $\lambda \in [0, \infty)$ . Then for any  $t \geq 0$  and any probability measures  $\mu, \nu$  on  $\mathbb{R}^d$ ,*

$$(3.12) \quad W_{f,1}(\mu p_t, \nu p_t) \leq \exp(-ct) W_{f,1}(\mu, \nu), \quad \text{and}$$

$$(3.13) \quad W_{\ell^1}(\mu p_t, \nu p_t) \leq A \exp(-ct) W_{\ell^1}(\mu, \nu),$$

where  $c = \min_{i=1, \dots, n} (c_i - 2\lambda \varphi_i(R_0^i)^{-1})$  and  $A = 2 \max_{i=1, \dots, n} \varphi_i(R_0^i)^{-1}$ .

The intuitive idea of proof for Theorem 3.1 is to construct a coupling  $(X_t, Y_t)$  of two solutions of (3.1) by applying a reflection coupling individually for each component  $(X_t^i, Y_t^i)$  if  $X_t^i \neq Y_t^i$ , and a basic coupling if  $X_t^i = Y_t^i$ . In the product case this just means that  $X_t^i = Y_t^i$  for any  $t \geq \tau^i$  where  $\tau^i = \inf\{t \geq 0 : X_t^i = Y_t^i\}$  is the coupling time for the  $i$ -th component. In the non-product case, however,  $X_t^i$  and  $Y_t^i$  can move apart again after the time  $\tau^i$  due to interactions with other components. In that case it is not clear how to define a coupling as described above rigorously. Instead we will use a regularized version where reflection coupling is applied to the  $i$ -th component whenever  $|X_t^i - Y_t^i| \geq \delta$  for a given constant  $\delta > 0$ , and basic coupling is applied whenever  $|X_t^i - Y_t^i| \leq \delta/2$ . A precise description of the coupling and the proofs of Theorem 3.1 and Corollary 3.3 are given in Sections 6 and 7 below.

**3.2. Consequences.** The contractivity results in Theorem 3.1 and Corollary 3.3 have corresponding consequences as the contractivity results in the non-product case, cf. Section 2.2 above. An important difference to be noted is, however, that on product spaces,

$$d_{f,w}(x, y) \leq \sum_{i=1}^n |x^i - y^i| \leq n^{1/2} |x - y|$$

by the Cauchy-Schwarz inequality. Therefore, an additional factor  $n$  occurs in the variance bounds from Corollaries 2.5, 2.6 and 2.7 on product spaces. Apart from this additional factor, all results in Section 2.2 carry over to the setup considered in Section 3.1.

**3.3. Interacting Langevin diffusions.** As an illustration of the results in Section 3.1, we consider a system

$$(3.14) \quad dX_t^i = -\frac{1}{2} \nabla U(X_t^i) dt - \sum_{j=1}^n a_{ij} \nabla V(X_t^i - X_t^j) dt + dB_t^i$$

of  $n$  interacting overdamped Langevin diffusions taking values in  $\mathbb{R}^k$  for some  $k \in \mathbb{N}$ . Here  $B^1, \dots, B^n$  are independent Brownian motions in  $\mathbb{R}^k$ ,  $U \in C^2(\mathbb{R}^k)$  is strictly convex outside a given ball, the interaction potential  $V$  is in  $C^2(\mathbb{R}^k)$  with bounded second derivatives, and  $a_{ij}$ ,  $1 \leq i, j \leq n$ , are finite real constants. For example, we are interested in *nearest-neighbour interactions* and *mean-field*



interactions given by

$$(3.15) \quad a_{ij} = \begin{cases} \alpha/2 & \text{if } i - j \equiv 1 \pmod{n} \text{ or } i - j \equiv -1 \pmod{n}, \\ 0 & \text{otherwise,} \end{cases}$$

$$(3.16) \quad a_{ij} = \alpha n^{-1} \quad \text{respectively,}$$

where  $\alpha \in \mathbb{R}$  is a finite coupling constant.

Choosing  $b_0^i(x^i) = -\nabla U(x^i)/2$  and  $\gamma^i(x) = \sum_{j=1}^n a_{ij} \nabla V(x^i - x^j)$ , we observe that the function

$$\kappa_i(r) = \inf \left\{ \int_0^1 \partial_{(x-y)/|x-y|}^2 U((1-t)x + ty) dt : x, y \in \mathbb{R}^k \text{ s.t. } |x - y| = r \right\}$$

does not depend on  $i$ . Let  $\varphi$  and  $f$  be the corresponding functions given by (2.6), and consider the distance

$$d_{1,f}(x, y) = \sum_{i=1}^n f(|x^i - y^i|).$$

Moreover, let  $c$  be given by (2.8) with  $\alpha = 1$ , i.e.,  $c$  is the lower bound for the contraction rate of the diffusion process  $Y$  in  $\mathbb{R}^k$  satisfying  $dY = -\frac{1}{2}\nabla U(Y) dt + dB$ . We note that  $\gamma$  satisfies the  $\ell^1$  Lipschitz condition (3.11) with

$$\lambda = M \cdot \max_i \sum_{j=1}^n (|a_{ij}| + |a_{ji}|)$$

where  $M = \sup \|\nabla^2 V\|$ . Therefore, if

$$\sum_{j=1}^n (|a_{ij}| + |a_{ji}|) \leq c \varphi(R_0) M^{-1}$$

then by Corollary 3.3, contractivity in the sense of (3.12) holds with contraction rate

$$\bar{c} = c - 2\lambda\varphi(R_0)^{-1} > 0.$$

In particular, in the nearest neighbour and mean field case, we obtain contractivity with a rate that does not depend on the dimension if  $\alpha$  is small:

**Corollary 3.4 (Mean field and nearest neighbour interactions).** *Let  $p_t$ ,  $t \geq 0$ , denote the transition kernels of the diffusion process on  $\mathbb{R}^{nk}$  solving (3.14). Suppose that  $\sup \|\nabla^2 V\| < \infty$  and that  $a_{ij}$  is given by (3.15) or by (3.16) with  $\alpha \in \mathbb{R}$ . Then there exist finite constants  $c, \theta, A \in (0, \infty)$  that do not depend on the dimension  $n$  such that*

$$(3.17) \quad W_{f,1}(\mu p_t, \nu p_t) \leq e^{(\theta\alpha - c)t} W_{f,1}(\mu, \nu), \quad \text{and}$$

$$(3.18) \quad W_{\ell^1}(\mu p_t, \nu p_t) \leq A e^{(\theta\alpha - c)t} W_{\ell^1}(\mu, \nu),$$

hold for any  $t \geq 0$  and any probability measures  $\mu, \nu$  on  $\mathbb{R}^{nk}$ . In particular, exponential contractivity holds for  $\alpha < c/\theta$ .

The bounds in (3.17) and (3.18) are not sharp. However, it is known that for example in mean field models where  $U$  is a double-well potential and  $V$  is quadratic, exponential contractivity with a rate independent of the dimension can not be expected to hold for large  $\alpha$ . Indeed, in this case the corresponding McKean-Vlasov process has several stationary distributions if  $\alpha > \alpha_1$  for some critical parameter  $\alpha_1 \in (0, \infty)$ , cf. [21, 22].

#### 4. PROOFS FOR REFLECTION COUPLING

In this section, we first motivate our particular choice of the function  $f$ , and we prove Theorem 2.2. Afterwards, we prove Corollaries 2.3, 2.5, 2.6 and 2.7.

Let  $r_t = \|X_t - Y_t\|$  where  $(X, Y)$  is a reflection coupling of two solutions of (1.1). Our goal is to find an explicit concave increasing function  $f : [0, \infty) \rightarrow [0, \infty)$  with  $f(0) = 0$  and  $f'(0) = 1$  such that  $e^{ct}f(r_t)$  is a (local) supermartingale for  $t$  less than the coupling time  $T$  with a constant  $c > 0$  that we are trying to maximize by the choice of  $f$ .

An application of Itô's formula to the s.d.e. (1.3) satisfied by the difference process  $Z_t = X_t - Y_t$  shows that the following Itô equations hold almost surely for  $t < T$  whenever  $f$  is  $C^1$  and  $f'$  is absolutely continuous:

$$\begin{aligned}
 d\|Z_t\|^2 &= 4|\sigma^{-1}Z_t|^{-1}\|Z_t\|^2 dW_t \\
 &\quad + 2Z_t \cdot G(b(X_t) - b(Y_t)) dt + 4|\sigma^{-1}Z_t|^{-2}\|Z_t\|^2 dt, \\
 dr_t &= 2|\sigma^{-1}Z_t|^{-1}r_t dW_t + r_t^{-1}Z_t \cdot G(b(X_t) - b(Y_t)) dt, \quad \text{and} \\
 df(r_t) &= 2|\sigma^{-1}Z_t|^{-1}r_t f'(r_t) dW_t \\
 (4.1) \quad &\quad + r_t^{-1}Z_t \cdot G(b(X_t) - b(Y_t))f'(r_t) dt + 2|\sigma^{-1}Z_t|^{-2}r_t^2 f''(r_t) dt.
 \end{aligned}$$

By definition of the function  $\kappa$ , the drift term on the right hand side of (4.1) is bounded from above by

$$(4.2) \quad \beta_t := 2|\sigma^{-1}Z_t|^{-2}r_t^2 \cdot \left( f''(r_t) - \frac{1}{4}r_t \kappa(r_t)f'(r_t) \right).$$

Hence the process  $e^{ct}f(r_t)$  is a supermartingale for  $t < T$  if  $\beta_t \leq -cf(r_t)$ . Since

$$(4.3) \quad |\sigma^{-1}z|^2 \leq \alpha\|z\|^2 \quad \text{for any } z \in \mathbb{R}^d,$$

a sufficient condition is

$$(4.4) \quad f''(r) - \frac{1}{4}r\kappa(r)f'(r) \leq -\frac{\alpha c}{2}f(r) \quad \text{for a.e. } r > 0.$$

We now first observe that this equation holds with  $c = 0$  (i.e.,  $f(r_t)$  is a supermartingale for  $t < T$ ) if  $f$  is chosen such that  $f'(r) = \varphi(r) = \exp(-\int_0^r s\kappa(s)ds/4)$ . Indeed,  $f(r) = \int_0^r \varphi(s)ds$  is the least concave among all concave functions  $f$  satisfying  $\beta_t \leq 0$ .

To satisfy the stronger condition  $\beta_t \leq -cf(r_t)$  with  $c < 0$ , we make the ansatz

$$(4.5) \quad f'(r) = \varphi(r)g(r)$$

with a decreasing absolutely continuous function  $g \geq 1/2$  such that  $g(0) = 1$ . Note that the condition  $g \geq 0$  is required to ensure that  $f$  is non-decreasing. By

replacing this condition by the stronger condition  $g \geq 1/2$ , we are loosing at most a factor 2 in the estimates below. On the other hand, the condition  $1/2 \leq g \leq 1$  has the huge advantage of ensuring that

$$(4.6) \quad \Phi/2 \leq f \leq \Phi$$

where  $\Phi(r) = \int_0^r \varphi(s) ds$ . The ansatz (4.5) yields

$$f'' = -\frac{1}{4} r \kappa^- f + \varphi g' \leq \frac{1}{4} r \kappa f + \varphi g',$$

i.e., Condition (4.4) is satisfied if

$$(4.7) \quad g' \leq -\frac{\alpha c}{2} f / \varphi. \quad \text{almost surely.}$$

We will see in the proof below that for  $r \geq R_1$ , Condition (4.4) is automatically satisfied since  $\kappa$  is sufficiently positive. Therefore, it is enough to assume that (4.7) holds on  $(0, R_1)$ .

Now on the one hand, if (4.7) is satisfied on  $(0, R_1)$  then

$$g(R_1) \leq 1 - \frac{\alpha c}{2} \int_0^{R_1} f(s) \varphi(s)^{-1} ds \leq 1 - \frac{\alpha c}{4} \int_0^{R_1} \Phi(s) \varphi(s)^{-1} ds.$$

This condition can only be satisfied with a function  $g$  taking values in  $[1/2, 1]$  if

$$\alpha c \leq 2 \int_0^{R_1} \Phi(s) \varphi(s)^{-1} ds.$$

On the other hand, by choosing

$$(4.8) \quad g'(r) = -\frac{\Phi(r)}{2\varphi(r)} \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} ds \quad \text{for } r < R_1,$$

Condition (4.7) is satisfied with the constant

$$\alpha c = 1 \int_0^{R_1} \Phi(s) \varphi(s)^{-1} ds.$$

This shows that up to a factor 2, choosing  $g$  as in (4.8) is the best we can do under the assumptions that we have made.

The considerations above explain the particular choice of the function  $f$  made in (2.6). Once this choice has been made, the proof of Theorem 2.2 is almost straightforward:

**Proof of Theorem 2.2.** As remarked above, the drift in the s.d.e. (4.1) for  $f(r_t)$  is bounded from above by  $\beta_t$  defined by (4.2). We now show that by our choice of  $f$  in (2.6), this expression is smaller than  $-cf(r_t)$  where  $c$  is given by

(2.8). Indeed, for  $r < R_1$ ,

$$\begin{aligned}
 (4.9) \quad f''(r) &= -\frac{1}{4}r\kappa(r)^-\varphi(r)g(r) - \frac{1}{2}\Phi(r) \Big/ \int_0^{R_1} \Phi(s)\varphi(s)^{-1} ds \\
 &\leq \frac{1}{4}r\kappa(r)f'(r) - \frac{1}{2}f(r) \Big/ \int_0^{R_1} \Phi(s)\varphi(s)^{-1} ds.
 \end{aligned}$$

For  $r > R_1$ , we have  $f'(r) = \varphi(r)/2 = \varphi(R_0)/2$  and  $\kappa(r)R_1(R_1 - R_0) \geq 8$  by definition of  $R_1$ , whence

$$\begin{aligned}
 (4.10) \quad f''(r) - \frac{1}{4}r\kappa(r)f'(r) &= -\frac{1}{8}r\kappa(r)\varphi(R_0) \leq -\frac{\varphi(R_0)}{R_1 - R_0} \cdot \frac{r}{R_1} \\
 &\leq -\frac{\varphi(R_0)}{R_1 - R_0} \cdot \frac{\Phi(r)}{\Phi(R_1)} \leq -\frac{1}{2}\Phi(r) \Big/ \int_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1} ds \\
 &\leq -\frac{1}{2}f(r) \Big/ \int_0^{R_1} \Phi(s)\varphi(s)^{-1} ds.
 \end{aligned}$$

Here we have used that for  $r \geq R_0$ , the function  $\varphi(r)$  is constant, and, therefore,  $\Phi(r) = \Phi(R_0) + (r - R_0)\varphi(R_0)$ , and

$$\begin{aligned}
 \int_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1} ds &= \int_{R_0}^{R_1} (\Phi(R_0) + (s - R_0)\varphi(R_0))\varphi(R_0)^{-1} ds \\
 &= \Phi(R_0)\varphi(R_0)^{-1}(R_1 - R_0) + (R_1 - R_0)^2/2 \\
 &\geq (R_1 - R_0)(\Phi(R_0) + (R_1 - R_0)\varphi(R_0))\varphi(R_0)^{-1}/2 \\
 &= (R_1 - R_0)\Phi(R_1)\varphi(R_0)^{-1}/2.
 \end{aligned}$$

By (4.9) and (4.10), we conclude that  $\beta_t \leq -cf(r_t)$ . Optional stopping in (4.1) at  $T_k = \inf\{t \geq 0 : r_t \notin (k^{-1}, k)\}$  now implies

$$\mathbb{E}[f(r_t) ; t < T_k] \leq -c \int_0^t \mathbb{E}[f(r_s) ; s < T_k] ds$$

for any  $k \in \mathbb{N}$  and  $t \geq 0$ . The assertion follows for  $k \rightarrow \infty$  since  $r_t = 0$  for  $t \geq T$ , and  $T = \sup T_k$  by non-explosiveness.  $\square$

**Proof of Corollary 2.3.** Let  $(X, Y)$  be a reflection coupling of two solutions of (1.1) with joint initial distribution  $(X_0, Y_0) \sim \eta$ . Then by Theorem 2.2,

$$\begin{aligned}
 W_f(\mu p_t, \nu p_t) &\leq \mathbb{E}[d_f(X_t, Y_t)] \leq e^{-ct} \mathbb{E}[d_f(X_0, Y_0)] \\
 &= e^{-ct} \int d_f(x, y) \eta(dx dy)
 \end{aligned}$$

for any  $t \geq 0$ . The estimate (2.9) now follows by taking the infimum over all couplings  $\eta$  of two given probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}^d$ . Moreover, (2.10) follows from (2.9) by (2.11).  $\square$

Next, we are going to prove the results in Section 2.2. Suppose that (2.12) holds,  $\|z\| = |\sigma^{-1}z|$  is the intrinsic metric, and  $b$  is in  $C^1$ . Corollary 2.3 implies

$$\int |y| p_t(x, dy) \leq \int |y| p_t(x_0, dy) + W^1(p_t(x, \cdot), p_t(x_0, \cdot)) < \infty$$

for any  $t \geq 0$  and any  $x \in \mathbb{R}^d$ . In particular,  $(p_t g)(x) = \int g(y) p_t(x, dy)$  is defined for any Lipschitz continuous function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ , and

$$|(p_t g)(x) - (p_t g)(y)| = |\mathbb{E}[g(X_t) - g(Y_t)]| \leq \|g\|_{\text{Lip}(f)} \mathbb{E}[d_f(X_t, Y_t)]$$

for any coupling  $(X_t, Y_t)$  of  $p_t(x, \cdot)$  and  $p_t(y, \cdot)$ . Hence by Theorem 2.2,

$$(4.11) \quad |(p_t g)(x) - (p_t g)(y)| \leq e^{-ct} \|g\|_{\text{Lip}(f)} d_f(x, y),$$

i.e.,  $p_t$  satisfies the exponential contractivity condition (2.13) w.r.t.  $\|\cdot\|_{\text{Lip}(f)}$ . If  $p_t g$  is  $C^1$  then by (4.11) and since

$$d_f(x, y) \leq \|x - y\| = |\sigma^{-1}(x - y)| \quad \forall x, y \in \mathbb{R}^d,$$

we obtain the uniform gradient bound

$$(4.12) \quad \sup |\sigma^T \nabla p_t g| \leq e^{-ct} \|g\|_{\text{Lip}(f)} \quad \forall t \geq 0.$$

It is well-known that this bound can be used to control variances w.r.t. the measures  $p_t(x, \cdot)$ :

**Lemma 4.1.** *For any  $t \geq 0$ ,  $x \in \mathbb{R}^d$ , and any Lipschitz continuous  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ ,*

$$(4.13) \quad \text{Var}_{p_t(x, \cdot)}(g) \leq \frac{1 - \exp(-2ct)}{2c} \|g\|_{\text{Lip}(f)}^2.$$

**Proof.** We may assume  $g \in C^2(\mathbb{R}^d)$  and  $t > 0$ . Then, by standard elliptic regularity results,  $(t, x) \mapsto (p_t g)(x)$  is differentiable in  $t$  and  $x$ , and

$$\frac{d}{dt} p_t g = \mathcal{L} p_t g = p_t \mathcal{L} g$$

where  $\mathcal{L} = \frac{1}{2} \sum a_{ij} \frac{\partial^2}{\partial x^i \partial x^j} + b(x) \cdot \nabla$ ,  $a = \sigma \sigma^T$ , is the generator of  $(X_t)$ , cf. e.g. [33, 31]. In particular, for  $s \in (0, t)$ ,

$$\begin{aligned} \frac{d}{ds} p_s (p_{t-s} g)^2 &= p_s (\mathcal{L}(p_{t-s} g)^2 - 2 p_{t-s} g \mathcal{L} p_{t-s} g) \\ &= p_s |\sigma^T \nabla p_{t-s} g|^2 \leq e^{-2c(t-s)} \|g\|_{\text{Lip}(f)}^2 \end{aligned}$$

by (4.12). Integrating w.r.t.  $s$ , we obtain

$$p_t g^2 - (p_t g)^2 \leq \frac{1 - \exp(-2ct)}{2c} \|g\|_{\text{Lip}(f)}^2,$$

which is equivalent to (4.13).  $\square$

By Lemma 4.1 and (4.11), we can now easily prove Corollaries 2.5, 2.6 and 2.7:

**Proof of Corollary 2.5.** Existence and uniqueness of a stationary distribution  $\mu$  for  $(p_t)_{t \geq 0}$  satisfying  $\int |y| \mu(dy) < \infty$  follows easily as in [27], Section 3: By

Corollary 2.3, the map  $\nu \mapsto \nu p_1$  is a contraction w.r.t. the distance  $W_f$  (equivalent to  $W^1$ ) on the complete metric space  $\mathcal{P}^1$  of all probability measures  $\nu$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  satisfying  $\int |y| \mu(dy) < \infty$ . Hence by the Banach fixed point theorem, there exists a unique probability measure  $\mu_0$  such that  $\mu_0 p_1 = \mu_0$ . It is then elementary to verify that the measure  $\mu = \int_0^1 \mu_0 p_s ds$  satisfies  $\mu p_t = \mu$  for any  $t \in [0, 1]$ , and hence for any  $t \in [0, \infty)$ . Moreover, by Corollary 2.3,

$$W_f(\mu, \nu p_t) = W_f(\mu p_t, \nu p_t) \leq e^{-ct} W_f(\mu, \nu)$$

for any  $\nu \in \mathcal{P}^1$ . In particular, as  $t \rightarrow \infty$ ,  $p_t(x, \cdot) \rightarrow \mu$  in  $\mathcal{P}^1$  for any  $x \in \mathbb{R}^d$ . The variance bound for  $\mu$  now follows from the corresponding bound for  $p_t(x, \cdot)$  in Lemma 4.1.  $\square$

**Proof of Corollary 2.6.** By Lemma 4.1,

$$\begin{aligned} \text{Cov}(g(X_t), h(X_{t+s})) &= \mathbb{E}[g(X_t) h(X_{t+s})] - \mathbb{E}[g(X_t)] \mathbb{E}[h(X_{t+s})] \\ &= \mathbb{E}[(g p_s h)(X_t)] - \mathbb{E}[g(X_t)] \mathbb{E}[(p_s h)(X_t)] = \text{Cov}_{p_t(x_0, \cdot)}(g, p_s h) \\ &\leq (1 - \exp(-2ct)) (2c)^{-1} \|g\|_{\text{Lip}(f)} \|p_s h\|_{\text{Lip}(f)} \end{aligned}$$

for any  $s, t \geq 0$ . The assertion now follows by (4.11).  $\square$

**Proof of Corollary 2.7.** The bound for the bias follows immediately from (4.11), since

$$\begin{aligned} \left| \mathbb{E} \left[ \frac{1}{t} \int_0^t g(X_s) ds - \int g d\mu \right] \right| &= \left| \frac{1}{t} \int_0^t \int (p_s g(x_0) - p_s g(y)) \mu(dy) ds \right| \\ &\leq \frac{1}{t} \int_0^t e^{-cs} ds \|g\|_{\text{Lip}(f)} \int d_f(x_0, y) \mu(dy). \end{aligned}$$

Moreover, by Corollary 2.6,

$$\begin{aligned} \text{Var} \left( \frac{1}{t} \int_0^t g(X_s) ds \right) &= \text{Cov} \left( \frac{1}{t} \int_0^t g(X_s) ds, \frac{1}{t} \int_0^t g(X_s) ds \right) \\ &= \frac{2}{t^2} \int_0^t \int_s^t \text{Cov}(g(X_s), g(X_u)) du ds \\ &\leq \frac{1}{ct^2} \int_0^t (1 - e^{-2cs}) \int_s^t e^{-c(u-s)} du ds \|g\|_{\text{Lip}(f)}^2 \\ &\leq \frac{1}{c^2 t} \|g\|_{\text{Lip}(f)}^2. \quad \square \end{aligned}$$

## 5. EXAMPLES

We now prove the results in Sections 2.3 and 2.4, including in particular Lemma 2.9, Lemma 2.12 and Theorem 2.14.

**Proof of Lemma 2.9 and Remark 2.10.** We first prove the lower bounds on the exponential decay rate  $c$  in (2.8) stated in (2.19), (2.17) and (2.18). Notice that

the constant  $c$  defined by (2.8) increases if  $\kappa(r)$  is replaced by a greater function. Indeed, for  $r \geq 0$ ,

$$(5.1) \quad \Phi(r)\varphi(r)^{-1} = \int_0^r \varphi(t)\varphi(r)^{-1} dt = \int_0^r \exp\left(\frac{1}{4} \int_t^r s\kappa(s)^- ds\right) dt,$$

whence  $R_0$ ,  $R_1$  and  $c^{-1} = \alpha \int_0^{R_1} \Phi(s)\varphi(s)^{-1} ds$  are decreasing functions of  $\kappa$ .

*Convex Case.* Suppose first that  $\kappa(r) \geq 0$  for any  $r \geq 0$  and  $\kappa(r) \geq K$  for  $r \geq R$  with constants  $K \in (0, \infty)$  and  $R \in [0, \infty)$ . Then  $R_0 = 0$ ,  $R_1 \leq \max(R, \sqrt{8/K})$ ,  $\varphi \equiv 1$ , and hence

$$c = (\alpha R_1^2/2)^{-1} \geq \alpha^{-1} \min(R^{-2}/2, K/4).$$

*Locally non-convex case.* Now suppose that  $\kappa(r) \geq -L$  for  $r \leq R$  and  $\kappa(r) \geq K$  for  $r > R$  with constants  $K, L \in (0, \infty)$  and  $R \in [0, \infty]$ . Since  $\varphi(r) = \varphi(R_0)$  and  $\Phi(r) = \Phi(R_0) + (r - R_0)\varphi(R_0)$  for  $r \geq R_0$ , we have

$$(5.2) \quad \begin{aligned} \alpha^{-1}c^{-1} &= \int_0^{R_1} \Phi(s)\varphi(s)^{-1} ds \\ &= \int_0^{R_0} \Phi(s)\varphi(s)^{-1} ds + (R_1 - R_0)\Phi(R_0)\varphi(R_0)^{-1} + (R_1 - R_0)^2/2. \end{aligned}$$

The lower curvature bounds imply the upper bounds

$$(5.3) \quad R_0 \leq R, \quad R_1 - R_0 \leq \min(8/(KR_0), \sqrt{8/K}), \quad \text{and}$$

$$(5.4) \quad \begin{aligned} \Phi(r)\varphi(r)^{-1} &\leq \int_0^r \exp(L(r^2 - t^2)/8) dt \\ &\leq \min(\sqrt{2\pi/L}, r) \exp(Lr^2/8) \quad \text{for } r \leq R_0. \end{aligned}$$

Since  $\exp x \leq 1 + (e - 1)x$  for  $x \in [0, 1]$  and

$$\int_0^x \exp(u^2) du \leq e + \int_1^x (2 - u^{-2}) \exp(u^2) du = x^{-1} \exp(x^2) \quad \text{for } x \geq 1,$$

we can conclude that

$$\begin{aligned} \int_0^{R_0} \Phi(r)\varphi(r)^{-1} dr &\leq \int_0^{R_0} r \exp(Lr^2/8) dr = 4L^{-1}(\exp(LR_0^2/8) - 1) \\ &\leq (e - 1)R_0^2/2 \quad \text{if } LR_0^2/8 \leq 1, \quad \text{and} \\ \int_0^{R_0} \Phi(r)\varphi(r)^{-1} dr &\leq \sqrt{\frac{2\pi}{L}} \int_0^{R_0} \exp\left(\frac{Lr^2}{8}\right) dr = \sqrt{\frac{8 \cdot 2\pi}{L^2}} \int_0^{\sqrt{LR_0^2/8}} \exp(u^2) du \\ &\leq 8\sqrt{2\pi}L^{-3/2}R_0^{-1} \exp(LR_0^2/8) \quad \text{if } LR_0^2/8 \geq 1. \end{aligned}$$



Combining these estimates, we obtain by (5.2), (5.3) and (5.4),

$$\begin{aligned} \alpha^{-1}c^{-1} &\leq (e-1)R^2/2 + e\sqrt{8/K}R + 4/K && \text{if } LR_0^2/8 \leq 1, \text{ and} \\ \alpha^{-1}c^{-1} &\leq 8\sqrt{2\pi}R^{-1}L^{-1/2}(L^{-1} + K^{-1})\exp(LR^2/8) + 32R^{-2}K^{-2} && \text{if } LR_0^2/8 \geq 1, \end{aligned}$$

where we have used that the function  $x \mapsto x^{-1}\exp(x^2)$  is increasing for  $x \geq 1$ .  $\square$

**Proofs for Example 2.11.** Consider the one-dimensional Langevin diffusion  $(X_t)$  with drift  $-\nabla U(x)/2$  and generator

$$(5.5) \quad \mathcal{L}v = \frac{1}{2}(v'' - U'v') = \frac{1}{2}e^U(e^{-U}v')'.$$

The assumption  $\liminf_{|x| \rightarrow \infty} U''(x) > 0$  implies that there is a unique strictly positive bounded eigenfunction  $v_1 \in C^2(0, \infty) \cap C([0, \infty))$  satisfying  $v_1(0) = 0$ ,  $v_1'(0) = 1$  and  $\mathcal{L}v_1 = -\lambda_1 v_1$ , where

$$\lambda_1 = \lambda_1(0, \infty) = \inf_{v \in C_0^\infty(0, \infty)} \frac{\frac{1}{2} \int_0^\infty v'(x)^2 \exp(-U(x)) dx}{\int_0^\infty v(x)^2 \exp(-U(x)) dx}$$

is the infimum of the spectrum of the self-adjoint realization of  $-\mathcal{L}$  with Dirichlet boundary conditions on  $(0, \infty)$ . Since  $\mathcal{L}v_1 = -\lambda_1 v_1$  and  $v_1$  is bounded, the process  $M_t = \exp(\lambda_1 t)v_1(X_t)$  is a martingale. Optional stopping applied to the diffusion with initial condition  $X_0 = x_0$  shows that

$$(5.6) \quad \begin{aligned} v_1(x_0) &= \mathbb{E}_{x_0}[M_0] = \mathbb{E}_{x_0}[M_{\tau_0 \wedge t}] = \mathbb{E}_{x_0}[\exp(\lambda_1 t)v_1(X_t); \tau_0 > t] \\ &\leq \exp(\lambda_1 t) \mathbb{P}_{x_0}[\tau_0 > t] \sup v_1 \end{aligned}$$

for any  $x_0 > 0$  and  $t \geq 0$ . Since  $v_1(x_0) > 0$  and  $\sup v_1 < \infty$ , the estimate (5.6) implies the asymptotic lower bound

$$(5.7) \quad \liminf_{t \rightarrow \infty} t^{-1} \log P_{x_0}[\tau_0 > t] \geq -\lambda_1(0, \infty).$$

Moreover, for any fixed  $t \leq \lambda^{-1}/4$ ,

$$\mathbb{P}_R[\tau_0 > t] \geq e^{-1/4} v_1(R) / \sup v_1 \geq 3/4$$

provided  $v_1(R) \geq \frac{3}{4}e^{1/4} \sup v_1 = 0.96 \dots \sup v_1$ . By the eigenfunction equation  $e^U(e^{-U}v_1')' = -\lambda_1 v_1$ , one verifies that the latter condition is satisfied whenever  $U$  is growing fast enough on  $[R, \infty)$ .

For bounding  $\lambda_1(0, \infty)$  from above let

$$v(x) = \min(\sqrt{L}x, 1) = \begin{cases} \sqrt{L}x & \text{if } x \leq 1/\sqrt{L}, \\ 1 & \text{if } x \geq 1/\sqrt{L}. \end{cases}$$

By the assumptions on  $U$ , the function  $v$  is contained in the weighted Sobolev space  $H_0^{1,2}((0, \infty), e^{-U} dx)$  (closure of  $C_0^\infty(0, \infty)$  w.r.t. the norm  $\|w\|^2 = \int_0^\infty (w^2 +$

$(w')^2) e^{-U} dx$ ). Therefore, if  $LR^2/4 \geq 1$  then (2.21) holds, since

$$\begin{aligned} \lambda_1 &\leq \frac{\frac{1}{2} \int v'(x)^2 \exp(-U(x)) dx}{\int v(x)^2 \exp(-U(x)) dx} \leq \frac{\int_0^{1/\sqrt{L}} L \exp(Lx^2/2) dx}{\int_0^{R/2} v(x)^2 \exp(Lx^2/2) dx} \\ &= \frac{L}{2} \frac{\int_0^1 \exp(y^2/2) dy}{\int_0^{\sqrt{LR^2/4}} \min(y, 1)^2 \exp(y^2/2) dy} \leq \frac{3Le^{1/2}}{2} \sqrt{\frac{LR^2}{4}} \exp\left(\frac{LR^2}{8}\right). \end{aligned}$$

Here we have used that by assumption,  $U(x) \geq -Lx^2/2$  for any  $x \in \mathbb{R}$  with equality for  $|x| < R/2$ , and for  $x \geq 1$ ,

$$\int_0^x \min(y, 1)^2 e^{y^2/2} dy = \int_0^1 \dots + \int_1^x \dots \geq \frac{1}{3} + \frac{1}{x} e^{x^2/2} - 1 \geq \frac{1}{3x} e^{x^2/2}$$

as  $(x^{-1} e^{x^2/2})' = (1 - x^{-2}) e^{x^2/2} \leq e^{x^2/2}$ .  $\square$

**Proof of Lemma 2.12.** Since  $b = b_0 + \gamma$ , we have

$$(x - y) \cdot G(b(x) - b(y)) = (x - y) \cdot G(b_0(x) - b_0(y)) + (x - y) \cdot G(\gamma(x) - \gamma(y))$$

for any  $x, y \in \mathbb{R}^d$ . Therefore, by (2.24) and by definition of  $\kappa$  and  $\kappa_0$ ,

$$(5.8) \quad \kappa(r)^- \leq \kappa_0(r)^- \quad \text{for any } r \leq R, \text{ and}$$

$$(5.9) \quad \kappa(r)^- \leq \kappa_0(r)^- + 4r^{-1} \sup \|\gamma\| \quad \text{for any } r \in (0, \infty).$$

In particular, if  $\gamma$  is bounded then  $\kappa$  satisfies the conditions in (2.3). Since the constant  $R_1(b)$  defined w.r.t.  $b$  is smaller than the corresponding constant  $R_1$  defined w.r.t.  $b_0$ , we obtain

$$\begin{aligned} \frac{1}{c} &\leq \int_0^{R_1} \int_0^s \exp\left(\frac{1}{4} \int_t^s u \kappa(u)^- du\right) dt ds \\ &\leq \int_0^{R_1} \int_0^s \exp\left(\frac{1}{4} \int_t^s u \kappa_0(u)^- du\right) \exp(R \sup \|\gamma\|) dt ds \\ &\leq \frac{1}{c_0} \cdot \exp(R \sup \|\gamma\|), \end{aligned}$$

i.e., (2.25) holds.

Similarly, if  $\gamma$  satisfies the one-sided Lipschitz condition (2.26) then

$$(5.10) \quad \kappa(r)^- \leq \kappa_0(r)^- + 2L \quad \text{for any } r \in (0, \infty).$$

Hence again the conditions in (2.3) are satisfied, and we obtain

$$\frac{1}{c} \leq \frac{1}{c_0} \cdot \exp\left(\frac{L}{2} \int_0^R r dr\right)$$

similarly as above, i.e., (2.27) holds.  $\square$

**Proof of Theorem 2.14.** Fix  $R > 0$  and probability measures  $\mu, \nu$  on  $\mathbb{R}^d$ . By definition of  $f_R$ ,

$$f_R''(r) \leq \frac{1}{4} r \kappa(r) f_R'(r) - f_R(r) \Big/ \int_0^R \frac{\Phi(s)}{\varphi(s)} ds$$

for any  $r < R$ . Therefore, similarly to the proof of Theorem 2.2, Equation (4.1) shows that the process  $e^{c_R t} f_R(r_t)$  is a local supermartingale for  $t < \hat{\tau}_R$  where

$$\hat{\tau}_R = \inf\{t \geq 0 : r_t > R\}.$$

Here  $r_t = \|X_t - Y_t\|$  again denotes the distance process for a reflection coupling  $(X_t, Y_t)$  of two solutions of (1.1) with initial distribution given by a coupling  $\eta$  of  $\mu$  and  $\nu$ . By optional stopping and Fatou's lemma, we thus obtain

$$\mathbb{E}[f_R(r_t); \hat{\tau}_R > t] \leq \mathbb{E}[f_R(r_{t \wedge \hat{\tau}_R})] \leq \exp(-c_R t) \mathbb{E}[f_R(r_0)]$$

for any  $t \geq 0$ , and hence

$$\begin{aligned} \mathbb{E}[f_R(r_t)] &\leq \exp(-c_R t) \mathbb{E}[f_R(r_0)] + \mathbb{P}[\hat{\tau}_R \leq t] \\ &\leq e^{-c_R t} \int f_R(\|x - y\|) \eta(dx dy) + \mathbb{P}_\mu[\tau_{R/2} \leq t] + \mathbb{P}_\nu[\tau_{R/2} \leq t]. \end{aligned}$$

The assertion now follows as in the proof of Corollary 2.3 by minimizing over all couplings  $\eta$  of  $\mu$  and  $\nu$ .  $\square$

## 6. COUPLINGS ON PRODUCT SPACES

Let  $d = \sum_{i=1}^n d_i$  with  $n, d_1, \dots, d_n \in \mathbb{N}$ . We now consider “componentwise” couplings for diffusion processes  $X_t = (X_t^1, \dots, X_t^n)$  and  $Y_t = (Y_t^1, \dots, Y_t^n)$  on  $\mathbb{R}^d$  satisfying the s.d.e.

$$(6.1) \quad dX_t^i = b^i(X_t) dt + dB_t^i, \quad i = 1, \dots, n,$$

with initial conditions  $X_0 \sim \mu$  and  $Y_0 \sim \nu$ . Here  $B^i, i = 1, \dots, n$ , are independent Brownian motions on  $\mathbb{R}^{d_i}$ , and  $b^i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d_i}$  are locally Lipschitz continuous functions such that the unique strong solution of (6.1) is non-explosive for any given initial condition.

Let  $\delta > 0$ . Suppose that  $\lambda^i, \pi^i : \mathbb{R}^d \rightarrow [0, 1], i = 1, \dots, n$ , are Lipschitz continuous functions such that

$$(6.2) \quad \lambda^i(z)^2 + \pi^i(z)^2 = 1 \quad \text{for any } z \in \mathbb{R}^d, \quad \text{and}$$

$$(6.3) \quad \lambda^i(z) = 0 \quad \text{if } |z^i| \leq \delta/2,$$

and let  $B^i$  and  $\tilde{B}^i, 1 \leq i \leq n$ , be independent Brownian motions on  $\mathbb{R}^{d_i}$ . Then a coupling of two solutions of (6.1) with initial distributions  $\mu$  and  $\nu$  is given by a strong solution of the system

$$(6.4) \quad \begin{aligned} dX_t^i &= b^i(X_t) dt + \lambda^i(Z_t) dB_t^i + \pi^i(Z_t) d\tilde{B}_t^i, \\ dY_t^i &= b^i(Y_t) dt + \lambda^i(Z_t) (I - 2e_t^i e_t^{i,T}) dB_t^i + \pi^i(Z_t) d\tilde{B}_t^i, \end{aligned}$$

$1 \leq i \leq n$ , with initial distribution  $(X_0, Y_0) \sim \eta$  where  $\eta$  is a coupling of  $\mu$  and  $\nu$ . Here we use the notation

$$Z_t = X_t - Y_t,$$

and  $e_t^i$  is a measurable process taking values in the unit sphere in  $\mathbb{R}^{d_i}$  such that

$$e_t^i = \begin{cases} Z_t^i / |Z_t^i| & \text{if } Z_t^i \neq 0, \\ u^i & \text{if } Z_t^i = 0, \end{cases}$$

where  $u^i$  is an arbitrary fixed unit vector in  $\mathbb{R}^{d_i}$ . Notice that by (6.3), the choice of  $u^i$  is not relevant for (6.4), which is a standard Itô s.d.e. in  $\mathbb{R}^{2d}$  with locally Lipschitz continuous coefficients. To see that (6.4) defines a coupling, we observe that  $(X_t)$  and  $(Y_t)$  satisfy (6.1) w.r.t. the processes  $\hat{B}_t = (\hat{B}_t^1, \dots, \hat{B}_t^n)$  and  $\check{B}_t = (\check{B}_t^1, \dots, \check{B}_t^n)$  defined by

$$\begin{aligned}\hat{B}_t^i &= \int_0^t \lambda^i(Z_s) dB_s^i + \int_0^t \pi^i(Z_s) d\tilde{B}_s^i, \\ \check{B}_t^i &= \int_0^t \lambda^i(Z_s) (I - 2e_s^i e_s^{i,T}) dB_s^i + \int_0^t \pi^i(Z_s) d\tilde{B}_s^i.\end{aligned}$$

By Lévy's characterization and by (6.2), both  $\hat{B}$  and  $\check{B}$  are indeed Brownian motions in  $\mathbb{R}^d$ , cp. the corresponding argument for reflection coupling.

**Remark 6.1.** (1) By Condition (6.3) and non-explosiveness of (6.1), the coupling process  $(X_t, Y_t)$  is defined for any  $t \geq 0$ .

(2) By choosing  $\lambda^i \equiv 0$  and  $\pi^i \equiv 1$  we recover the *basic coupling*, i.e., the same noise is applied to both processes  $X$  and  $Y$ .

(3) A *componentwise reflection coupling* would be informally given by choosing  $\lambda^i(z) = 1$  if  $z^i \neq 0$  and  $\lambda^i(z) = 0$  if  $z^i = 0$ . As this function is not continuous and  $e^i(z) = z^i/|z^i|$  also has a discontinuity at zero, it is not obvious how to make sense of this coupling rigorously. Instead, we will use below an approximate componentwise reflection coupling where  $\lambda^i(z) = 1$  if  $|z^i| \geq \delta$  and  $\lambda^i(z) = 0$  if  $|z^i| \leq \delta/2$  for a small positive constant  $\delta$ .

By subtracting the equations for  $X$  and  $Y$  in (6.4), we see that the difference process  $Z = X - Y$  satisfies the s.d.e.

$$(6.5) \quad dZ_t^i = (b^i(X_t) - b^i(Y_t)) dt + 2\lambda^i(Z_t) e_t^i dW_t^i,$$

$i = 1, \dots, n$ , where the processes

$$W_t^i = \int_0^t e_s^{i,T} dB_s^i, \quad 1 \leq i \leq n,$$

are independent *one-dimensional* Brownian motions.

Let  $r_t^i = |X_t^i - Y_t^i|$  denote the Euclidean norm of  $Z_t^i$ . The next lemma is crucial for quantifying contraction properties of the coupling given by (6.4):

**Lemma 6.2.** *Suppose that  $f : [0, \infty) \rightarrow [0, \infty)$  is a strictly increasing concave function in  $C^1([0, \infty))$  such that  $f'$  is absolutely continuous on  $(0, \infty)$ . Then for any  $i = 1, \dots, n$ , the process  $f(r_t^i)$  satisfies the Itô equation*

$$\begin{aligned}(6.6) \quad f(r_t^i) &= f(r_0^i) + 2 \int_0^t \lambda^i(X_s - Y_s) f'(r_s^i) dW_s^i \\ &+ \int_0^t \{e_s^i \cdot (b^i(X_s) - b^i(Y_s)) f'(r_s^i) + 2\lambda^i(X_s - Y_s)^2 f''(r_s^i)\} ds.\end{aligned}$$

**Remark 6.3.** The lemma shows in particular that the process  $r_t^i$  satisfies

$$(6.7) \quad dr_t^i = e_t^i \cdot (b^i(X_t) - b^i(Y_t)) dt + 2\lambda^i(X_t - Y_t) dW_t^i.$$

Notice that in this equation, the drift term does not depend on the choice of  $\lambda$ .

**Proof of Lemma 6.2.** Recall that  $e_t^i = Z_t^i/|Z_t^i|$  if  $r_t^i = |Z_t^i| \neq 0$ . Since the function  $y \mapsto y/|y|$  is smooth on  $\mathbb{R}^{d_i} \setminus \{0\}$  and  $x \mapsto \sqrt{x}$  is smooth on  $(0, \infty)$ , we can apply Itô's formula and (6.5) to show that the Itô equations

$$(6.8) \quad \begin{aligned} d|Z^i|^2 &= 2Z^i \cdot (b^i(X) - b^i(Y)) dt + 4\lambda^i(Z)^2 dt + 4\lambda^i(Z) |Z^i| dW^i, \\ dr^i &= \frac{1}{2r^i} d|Z^i|^2 - \frac{1}{8(r^i)^3} d[|Z^i|^2] \\ &= e^i \cdot (b^i(X) - b^i(Y)) dt + 2\lambda^i(X - Y) dW^i \end{aligned}$$

hold almost surely on any stochastic interval  $[\tau_1, \tau_2]$  such that  $Z_t^i \neq 0$  a.s. for  $\tau_1 \leq t \leq \tau_2$ .

On the other hand, suppose that  $|Z^i| < \delta/2$  a.s. on a stochastic interval  $[\tau_3, \tau_4]$ . Then on  $[\tau_3, \tau_4]$ ,  $\lambda(Z) \equiv 0$  by (6.3), and hence  $Z^i$  is almost surely absolutely continuous with

$$dZ^i/dt = b^i(X) - b^i(Y) \quad \text{a.e. on } [\tau_3, \tau_4].$$

This implies that  $r^i = |Z^i|$  is almost surely absolutely continuous on  $[\tau_3, \tau_4]$  as well with

$$(6.9) \quad dr^i/dt = e^i \cdot (b^i(X) - b^i(Y)) \quad \text{a.e. on } [\tau_3, \tau_4],$$

which is equivalent to (6.7) on  $[\tau_3, \tau_4]$ . Note that the value of  $e^i$  for  $Z^i = 0$  is not relevant here, since  $Z^i$  can only stay at 0 for a positive amount of time if  $b^i(X) - b^i(Y)$  vanishes during that time interval.

Since  $\mathbb{R}_+$  is the union of countably many stochastic intervals of the first and second type considered above, the Itô equation (6.7) holds almost surely on  $\mathbb{R}_+$ . The assertion (6.6) now follows from (6.7) by another application of Itô's formula. Here it is enough to assume that  $f$  is  $C^1$  on  $[0, \infty)$  and  $f'$  is absolutely continuous on  $(0, \infty)$  because  $\lambda^i(X_s - Y_s)$  vanishes for  $r_s^i < \delta/2$ .  $\square$

We now fix weights  $w_1, \dots, w_n \in [0, \infty)$  and strictly increasing concave functions  $f_1, \dots, f_n \in C^1([0, \infty)) \cap C^2((0, \infty))$  such that  $f_i(0) = 0$  for any  $i$ . Consider

$$(6.10) \quad \rho_t = \sum_{i=1}^n f_i(r_t^i) w_i = d_{f,w}(X_t, Y_t)$$

where  $d_{f,w}$  is defined by (3.3). By Lemma 6.2,

$$(6.11) \quad \begin{aligned} d\rho_t &= \sum_{i=1}^n (e_t^i \cdot (b^i(X_t) - b^i(Y_t)) f_i'(r_t^i) + 2\lambda^i(X_t - Y_t)^2 f_i''(r_t^i)) w_i dt \\ &\quad + 2 \sum_{i=1}^n \lambda^i(X_t - Y_t) f_i'(r_t^i) dW_t^i. \end{aligned}$$

Notice that the last term on the right hand side is a martingale since  $\lambda^i$  and  $f'_i$  are bounded. This enables us to control the expectation  $\mathbb{E}[\rho_t]$  if we can bound the drift in (6.11) by  $m - c\rho_t$  for constants  $m, c \in (0, \infty)$ :

**Lemma 6.4.** *Let  $m, c \in (0, \infty)$  and suppose that*

$$(6.12) \quad \sum_{i=1}^n \left( cf_i(r^i) + (x^i - y^i) \cdot (b^i(x) - b^i(y)) \frac{f'_i(r^i)}{r^i} + 2\lambda^i(x - y)^2 f''_i(r^i) \right) w_i \leq m$$

*holds for any  $x, y \in \mathbb{R}^d$  with  $r^i := |x^i - y^i| > 0 \ \forall i \in \{1, \dots, n\}$ . Then*

$$(6.13) \quad \mathbb{E}[\rho_t] \leq e^{-ct} \mathbb{E}[\rho_0] + m(1 - e^{-ct})/c \quad \text{for any } t \geq 0.$$

**Proof.** We first note that by continuity of  $b^i$  and  $f'_i$ , (6.12) implies that

$$(6.14) \quad \sum_{i=1}^n (cf_i(r^i) + e^i \cdot (b^i(x) - b^i(y)) f'_i(r^i) + 2\lambda^i(x - y)^2 f''_i(r^i)) w_i \leq m$$

holds for any  $x, y \in \mathbb{R}^d$  (even if  $x^i - y^i = 0$ ) provided  $e^i = (x^i - y^i)/r^i$  if  $r^i > 0$  and  $e^i$  is an arbitrary unit vector if  $r^i = 0$ . Indeed, we obtain (6.14) by applying (6.12) with  $x^i$  replaced by  $x^i + he^i$  whenever  $x^i - y^i = 0$  and taking the limit as  $h \downarrow 0$ . In particular, by (6.14), the drift term  $\beta_t$  in (6.11) is bounded from above by

$$\beta_t \leq m - \sum_{i=1}^n cf_i(r_t^i) w_i = m - c\rho_t.$$

Therefore by (6.11) and by the Itô product rule,

$$d(e^{ct}\rho) = e^{ct}d\rho + ce^{ct}\rho dt \leq e^{ct}m dt + dM$$

where  $M$  is a martingale, and thus

$$\mathbb{E}[e^{ct}\rho_t] \leq \mathbb{E}[\rho_0] + m \int_0^t e^{cs} ds \quad \text{for any } t \geq 0.$$

□

Since  $f''_i \leq 0$ , the process  $\rho_t$  is decreasing more rapidly (or growing more slowly) if  $\lambda^i$  takes larger values. In particular, the decay properties of  $\rho_t$  would be optimized when  $\lambda^i(z) = 1$  for any  $z$  with  $z^i \neq 0$ . This optimal choice of  $\lambda^1, \dots, \lambda^n$  would correspond to a componentwise reflection coupling, but it violates Condition (6.3). It is perhaps possible to construct a corresponding coupling process by an approximation argument. For our purpose of bounding the Kantorovich distance  $W_{f,w}(\mu p_t, \nu p_t)$  this is not necessary. Indeed, it will be sufficient to consider approximate componentwise reflection couplings where (6.2) and (6.3) are satisfied and  $\lambda^i(z) = 1$  whenever  $|z^i| > \delta$ . The limit  $\delta \downarrow 0$  will then be considered for the resulting estimates of the Kantorovich distance but not for the coupling processes.

## 7. APPLICATION TO INTERACTING DIFFUSIONS

We will now apply the couplings introduced in Section 6 to prove the contraction properties for systems of interacting diffusions stated in Theorem 3.1 and Corollary 3.3. We consider the setup described in Section 3.1, i.e.,

$$(7.1) \quad b^i(x) = b_0^i(x^i) + \gamma^i(x) \quad \text{for } i = 1, \dots, n$$

with  $b_0^i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}^{d_i}$  locally Lipschitz such that  $\kappa_i$  defined by (3.4) is continuous on  $(0, \infty)$  with

$$(7.2) \quad \liminf_{r \rightarrow \infty} \kappa_i(r) > 0 \quad \text{and} \quad \lim_{r \rightarrow 0} r \kappa_i(r) = 0 \quad \text{for any } 1 \leq i \leq n.$$

The functions  $f_i$  are defined via  $\kappa_i$ , and  $c_i$  is the corresponding contraction rate given by (3.7).

**Proof of Theorem 3.1.** We fix  $\delta > 0$  and Lipschitz continuous functions  $\lambda^i, \mu^i : \mathbb{R}^d \rightarrow [0, 1]$ ,  $1 \leq i \leq n$ , such that (6.2) and (6.3) hold and  $\lambda^i(z) = 1$  if  $|z^i| \geq \delta$ . Let  $(X_t, Y_t)$  denote a corresponding approximate componentwise reflection coupling of two solutions of (3.1) given by (6.4), and let  $\rho_t = d_{f,w}(X_t, Y_t)$ . We will apply Lemma 6.4 which requires bounding the right hand side in (6.12). For this purpose recall that  $f_i$  and  $c_i$  have been chosen in such a way that

$$2f_i''(r) - \frac{1}{2}r\kappa_i(r)f_i'(r) \leq -c_i f_i(r) \quad \forall r > 0,$$

cf. (4.9) and (4.10). Therefore, by (7.1) and by definition of  $\kappa_i$ ,

$$(7.3) \quad \begin{aligned} & (x^i - y^i) \cdot (b^i(x) - b^i(y)) f_i'(r^i)/r^i + 2\lambda^i(x - y)^2 f_i''(r^i) \\ & \leq -\frac{1}{2}r^i \kappa_i(r^i) f_i'(r^i) + |\gamma^i(x) - \gamma^i(y)| f_i'(r^i) + 2\lambda^i(x - y)^2 f_i''(r^i) \\ & \leq -\lambda^i(x - y)^2 c_i f_i(r^i) + |\gamma^i(x) - \gamma^i(y)| - \frac{1}{2}(1 - \lambda^i(x - y)^2) r^i \kappa_i(r^i) f_i'(r^i) \\ & \leq -c_i f_i(r^i) + |\gamma^i(x) - \gamma^i(y)| + c_i \delta + \frac{1}{2} \sup_{r < \delta} (r \kappa_i(r)^-) \end{aligned}$$

for any  $x, y \in \mathbb{R}^d$  with  $r^i = |x^i - y^i| > 0$ . Here we have used that  $0 \leq f_i' \leq 1$ , and that  $\lambda^i(x - y) \neq 1$  only if  $r^i < \delta$ . In this case,  $f_i(r^i) \leq r^i \leq \delta$ . By (7.3) and by the assumption (3.8) on  $\gamma^i$ , we obtain

$$\begin{aligned} & \sum_{i=1}^n ((x^i - y^i) \cdot (b^i(x) - b^i(y)) f_i'(r^i)/r^i + 2\lambda^i(x - y)^2 f_i''(r^i)) w_i \\ & \leq m(\delta) + \sum_{i=1}^n (-c_i + \varepsilon_i) f_i(r^i) w_i \leq m(\delta) - c \sum_{i=1}^n f_i(r^i) w_i \end{aligned}$$

for  $x, y$  as above, where

$$m(\delta) = \sum_{i=1}^n (c_i \delta + \frac{1}{2} \sup_{r < \delta} (r \kappa_i(r)^-))$$



is a finite constant by (7.2), and  $c = \min_{i=1,\dots,n}(c_i - \varepsilon_i)$ . Hence (6.12) is satisfied with  $c$  and  $m(\delta)$  and, therefore,

$$(7.4) \quad \mathbb{E}[\rho_t] \leq e^{-ct} \mathbb{E}[\rho_0] + m(\delta)(1 - e^{-ct})/c.$$

By choosing the coupling process  $(X_t, Y_t)$  with initial distribution given by a coupling  $\eta$  of probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}^d$ , we conclude that

$$(7.5) \quad \begin{aligned} W_{f,w}(\mu p_t, \nu p_t) &\leq \mathbb{E}[d_{f,w}(X_t, Y_t)] = \mathbb{E}[\rho_t] \\ &\leq e^{-ct} \int d_{f,w}(x, y) \eta(dx dy) + m(\delta)(1 - e^{-ct})/c \end{aligned}$$

for any  $t \geq 0$ . Moreover, by (3.6),  $m(\delta) \rightarrow 0$  as  $\delta \downarrow 0$ . Hence the assertion (3.9) follows from (7.5) by taking the limit as  $\delta \downarrow 0$  and minimizing over all couplings  $\eta$  of  $\mu$  and  $\nu$ . Finally, (3.10) follows from (3.9) since  $\varphi(R_0^i)r/2 \leq f_i(r) \leq r$  implies

$$A^{-1} d_{\ell^1}(x, y) \leq d_{f,w}(x, y) = \sum f_i(|x^i - y^i|) w_i \leq d_{\ell^1}(x, y). \quad \square$$

**Proof of Corollary 3.3.** The  $\ell^1$ -Lipschitz condition (3.11) for  $\gamma$  implies that (3.8) holds with  $w_i = 1$  for any  $i$ , and

$$\lambda \varepsilon_i^{-1} = \inf_{r>0} f_i(r) = f'_i(R_1^i) = \varphi_i(R_0^i)/2,$$

i.e.,  $\varepsilon_i = 2\lambda/\varphi_i(R_0^i)$ . The assertion now follows from Theorem 3.1. □

## REFERENCES

1. Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000, With a preface by Dominique Bakry and Michel Ledoux. MR 1845806 (2002g:46132)
2. D. Bakry and Michel Émery, *Diffusions hypercontractives*, Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, pp. 177–206. MR 889476 (88j:60131)
3. Dominique Bakry, *L'hypercontractivité et son utilisation en théorie des semigroupes*, Lectures on probability theory (Saint-Flour, 1992), Lecture Notes in Math., vol. 1581, Springer, Berlin, 1994, pp. 1–114.
4. Dominique Bakry, Patrick Cattiaux, and Arnaud Guillin, *Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré*, J. Funct. Anal. **254** (2008), no. 3, 727–759. MR 2381160 (2009g:60104)
5. François Bolley, Ivan Gentil, and Arnaud Guillin, *Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations*, J. Funct. Anal. **263** (2012), no. 8, 2430–2457. MR 2964689
6. N. Bou-Rabee, M. Hairer, and E. Vanden-Eijnden, *Non-asymptotic mixing of the MALA algorithm*, ArXiv e-prints (2010).
7. Nawaf Bou-Rabee and Eric Vanden-Eijnden, *A patch that imparts unconditional stability to explicit integrators for Langevin-like equations*, J. Comput. Phys. **231** (2012), no. 6, 2565–2580. MR 2881033
8. P. Cattiaux and A. Guillin, *Semi Log-Concave Markov Diffusions*, ArXiv e-prints (2013).
9. Guan-Yu Chen and Laurent Saloff-Coste, *The cutoff phenomenon for ergodic Markov processes*, Electron. J. Probab. **13** (2008), no. 3, 26–78. MR 2375599 (2009b:60221)

10. Mu Fa Chen, *From Markov chains to nonequilibrium particle systems*, World Scientific Publishing Co. Inc., River Edge, NJ, 1992. MR 1168209 (94a:60135)
11. Mu Fa Chen and Shao Fu Li, *Coupling methods for multidimensional diffusion processes*, Ann. Probab. **17** (1989), no. 1, 151–177. MR 972776 (90a:60134)
12. Mu Fa Chen and Feng Yu Wang, *Estimation of the first eigenvalue of second order elliptic operators*, J. Funct. Anal. **131** (1995), no. 2, 345–363. MR 1345035 (96g:35142)
13. Mu-Fa Chen and Feng-Yu Wang, *Estimation of spectral gap for elliptic operators*, Trans. Amer. Math. Soc. **349** (1997), no. 3, 1239–1267. MR 1401516 (97h:35175)
14. M. Cranston, *Gradient estimates on manifolds using coupling*, J. Funct. Anal. **99** (1991), no. 1, 110–124. MR 1120916 (93a:58175)
15. Persi Diaconis, *The cutoff phenomenon in finite Markov chains*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), no. 4, 1659–1664. MR 1374011 (97b:60112)
16. Persi Diaconis and Laurent Saloff-Coste, *Separation cut-offs for birth and death chains*, Ann. Appl. Probab. **16** (2006), no. 4, 2098–2122. MR 2288715 (2008g:60226)
17. Andreas Eberle, *Reflection coupling and Wasserstein contractivity without convexity*, C. R. Math. Acad. Sci. Paris **349** (2011), no. 19–20, 1101–1104. MR 2843007 (2012h:60242)
18. Andreas Eberle, *Metropolis-Hastings algorithms for perturbations of Gaussian measures in high dimensions: Contraction properties and error bounds in the logconcave case*, to appear in Annals of Applied Probab. (2013).
19. Mark Freidlin, *Functional integration and partial differential equations*, Annals of Mathematics Studies, vol. 109, Princeton University Press, Princeton, NJ, 1985. MR 833742 (87g:60066)
20. Martin Hairer and Jonathan C. Mattingly, *Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations*, Ann. Probab. **36** (2008), no. 6, 2050–2091. MR 2478676 (2010i:35295)
21. S. Herrmann and J. Tugaut, *Non-uniqueness of stationary measures for self-stabilizing processes*, Stochastic Process. Appl. **120** (2010), no. 7, 1215–1246. MR 2639745 (2011i:60102)
22. Samuel Herrmann and Julian Tugaut, *Self-stabilizing processes: uniqueness problem for stationary measures and convergence rate in the small-noise limit*, ESAIM Probab. Stat. **16** (2012), 277–305. MR 2956576
23. Elton P. Hsu, *Stochastic analysis on manifolds*, Graduate Studies in Mathematics, vol. 38, American Mathematical Society, Providence, RI, 2002. MR 1882015 (2003c:58026)
24. Aldéric Joulin, *A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature*, Bernoulli **15** (2009), no. 2, 532–549. MR 2543873 (2010j:60213)
25. Aldéric Joulin and Yann Ollivier, *Curvature, concentration and error estimates for Markov chain Monte Carlo*, Ann. Probab. **38** (2010), no. 6, 2418–2442. MR 2683634 (2011j:60229)
26. Wilfrid S. Kendall, *Coupled Brownian motions and partial domain monotonicity for the Neumann heat kernel*, J. Funct. Anal. **86** (1989), no. 2, 226–236. MR 1021137 (90m:58203)
27. Tomasz Komorowski and Anna Walczuk, *Central limit theorem for Markov processes with spectral gap in the Wasserstein metric*, Stochastic Process. Appl. **122** (2012), no. 5, 2155–2184. MR 2921976
28. David A. Levin, Malwina J. Luczak, and Yuval Peres, *Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability*, Probab. Theory Related Fields **146** (2010), no. 1–2, 223–265. MR 2550363 (2011d:82062)
29. Torgny Lindvall, *Lectures on the coupling method*, Dover Publications Inc., Mineola, NY, 2002, Corrected reprint of the 1992 original. MR 1924231
30. Torgny Lindvall and L. C. G. Rogers, *Coupling of multidimensional diffusions by reflection*, Ann. Probab. **14** (1986), no. 3, 860–872. MR 841588 (88b:60179)
31. Gilles Royer, *Une initiation aux inégalités de Sobolev logarithmiques*, Cours Spécialisés [Specialized Courses], vol. 5, Société Mathématique de France, Paris, 1999. MR 1704288 (2000k:60002)

- 32. ———, *An initiation to logarithmic Sobolev inequalities*, SMF/AMS Texts and Monographs, vol. 14, American Mathematical Society, Providence, RI, 2007, Translated from the 1999 French original by Donald Babbitt.
- 33. Daniel W. Stroock, *Partial differential equations for probabilists*, Cambridge Studies in Advanced Mathematics, vol. 112, Cambridge University Press, Cambridge, 2012, Paperback edition of the 2008 original. MR 2932517
- 34. Hermann Thorisson, *Coupling, stationarity, and regeneration*, Probability and its Applications (New York), Springer-Verlag, New York, 2000. MR 1741181 (2001b:60003)
- 35. Max-K. von Renesse and Karl-Theodor Sturm, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), no. 7, 923–940. MR 2142879 (2006j:53048)
- 36. Feng-Yu Wang, *Functional inequalities, Markov semigroups and spectral theory*, Science Press, 2005.

UNIVERSITÄT BONN, INSTITUT FÜR ANGEWANDTE MATHEMATIK, ENDENICHER ALLEE 60,  
53115 BONN, GERMANY

*E-mail address:* eberle@uni-bonn.de

*URL:* <http://www.uni-bonn.de/~eberle>